

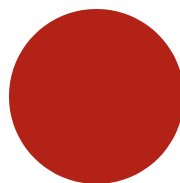
MESTRADO EM CIÊNCIA DA INFORMAÇÃO

REPOSITÓRIOS E PRESERVAÇÃO DIGITAL: proposta de requisitos para a integração do RI UFRN com a Rede Cariniana

Elisângela Alves de Moura

M
2015

UNIDADES ORGÂNICAS ENVOLVIDAS
FACULDADE DE ENGENHARIA
FACULDADE DE LETRAS



Elisângela Alves de Moura

Repositórios e preservação digital:
Proposta de requisitos para a integração do RI UFRN
com a Rede Cariniana

Dissertação realizada no âmbito do Mestrado em Ciência da Informação, orientada
pela Professora Doutora Maria Cristina de Carvalho Alves Ribeiro e co-orientada pelo
Doutor Miguel Ángel Márdero Arellano

Faculdade de Engenharia e Faculdade de Letras
Universidade do Porto

Julho de 2015

Repositórios e preservação digital: proposta de requisitos para a integração do RI UFRN com a Rede Cariniana

Elisângela Alves de Moura

Dissertação realizada no âmbito do Mestrado em Ciência da Informação, orientada
pela Professora Doutora Maria Cristina de Carvalho Alves Ribeiro e co-orientada pelo
Doutor Miguel Ángel Márdero Arellano

Membros do Júri

Professor Doutor António Manuel Lucas Soares- Presidente
Faculdade de Engenharia - Universidade do Porto

Professora Doutora Ana Alice Rodrigues Pereira Baptista - Arguente
Escola de Engenharia - Universidade do Minho

Professora Doutora Maria Cristina de Carvalho Alves Ribeiro - Orientadora
Faculdade de Engenharia - Universidade do Porto

Agradecimentos

À minha orientadora Doutora Cristina Ribeiro, pela orientação e apoio.

Ao meu co-orientador Miguel Ángel Mardero Arelano, muito obrigada por ter aceito colaborar com este estudo.

Aos professores do MCI.

Aos dirigentes da UFRN e BCZM, que concederam o meu afastamento das atividades laborais, o que me permitiu viver essa experiência.

À equipe da Rede Cariniana, muito obrigada pela colaboração.

Aos amigos e colegas de trabalho do Setor de Repositórios Digitais da BCZM, muito obrigada pela compreensão e apoio.

Aos colegas de mestrado que compartilharam comigo as alegrias e angústias desse percurso.

Às colegas do MCI e amigas na vida, Rúbia e Ângela, pela amizade, apoio, companheirismo, risadas e comilanças. Mas ‘bah’ como rimos, e sobretudo de nós mesmas. Sem vocês provavelmente teria conseguido, mas “Rais parta” se teria valido tanto a pena!

Às amigas brasileiras que durante um ano me apoiaram a distância e que no ano seguinte me apoiaram ainda mais, entendendo meu “desaparecimento” mesmo quando estava ao lado, e que estiveram sempre prontas a ajudar todas as vezes que precisei!

À minha família, que sempre me apoiou, independentemente das minhas decisões. Especialmente aos meus pais, que apesar de terem “poucos estudos” fizeram o possível para que nós, eu e meus irmãos, tivéssemos melhores oportunidades e nos ensinaram o quão importante é aprender, sempre. E à eles dedico este estudo!

Aos que contribuíram de alguma forma, com um abraço, um “fino”, café, vinho, ou mesmo críticas.

A todos a minha gratidão!

Resumo

Os repositórios institucionais compreendem uma infraestrutura tecnológica e de serviço que tem como finalidade a publicação, acesso, uso e armazenamento de informação. Esses repositórios lidam com a produção intelectual de uma determinada instituição, e estão diretamente ligados as instituições de ensino superior. Porém, o ambiente tecnológico no qual se inserem é frágil, e expõe os formatos e suportes à obsolescência. Com isso, a preservação digital, enquanto conjunto de processos que visa garantir a salvaguarda e acesso de longo prazo a essa produção, ganha fundamental importância. Nesse sentido, atua a rede brasileira de serviços de preservação intitulada Rede Cariniana, gerida pelo IBICT. A Rede utiliza o modelo de preservação digital distribuída, operacionalizada através do *software* LOCKSS. Com o objetivo de apoiar um acordo de cooperação entre a UFRN e o IBICT, este estudo propôs um conjunto de requisitos mínimos, direcionado para a UFRN e seu repositório institucional, com o intuito de adequá-lo à Rede Cariniana, para com isso possibilitar a preservação e o acesso de longo prazo à produção da UFRN. O conjunto é voltado principalmente para a preparação dos documentos, visando tratá-los, para que sejam inseridos de forma adequada no repositório e posteriormente incorporados à Rede Cariniana. Com o conjunto proposto pretende-se apoiar o acordo de cooperação e tratar a preservação digital da produção intelectual da UFRN a curto prazo, para que a instituição possa construir, a médio e longo-prazo, um Plano de Preservação Digital mais ambicioso, ou mesmo um Plano para certificação de seu repositório.

Palavras-chave: Repositório digital, preservação digital, repositório institucional da UFRN, Rede Cariniana.

Abstract

Institutional repositories include a technological and service infrastructure that aims to publication, access, use and store information. These repositories handle intellectual production of a given institution, and are directly connected with higher education institutions. However, their technological environment is fragile, and exposes the formats and supports to obsolescence. Therefore, digital preservation, as a set of processes aimed at ensuring the safety and long-term access to this production, gain fundamental importance. “Rede Cariniana”, the Brazilian network of preservation services managed by IBICT, is committed to solve this problem. The network uses the model of distributed digital preservation, operationalized through the LOCKSS *software*. With the goal of supporting a cooperation agreement between the UFRN and IBICT, this study proposes a set of minimum requirements, directed to the UFRN and its institutional repository, to tailor the repository to “Rede Cariniana”, enabling the preservation and long-term access to the intellectual production of UFRN. The requirements concern mainly the preparation of documents, so that they are deposited properly in the repository and subsequently incorporated into “Rede Cariniana”. The proposed set of requirements is aimed at supporting the working agreement and treat the digital preservation of UFRN intellectual production in the short term, enabling the institution to build, in a medium and long term, a more ambitious digital preservation plan, or a certification plan for its repository.

Keywords: Digital Repository, digital preservation, UFRN institutional repository, Cariniana Network.

Lista de Figuras e Tabelas

Figura 1 – Modelo de referência OAIS	28
Figura 2 – Modelo de informação OAIS	29
Figura 3 – Funcionamento básico do OAI-PMH	44
Figura 4 – Funcionamento do DSpace	49
Figura 5 – Organização do RI UFRN	55
Figura 6 – Comunidade BDTD	56
Figura 7 – Depósito de Itens	58
Figura 8 – Fluxo de aprovação	59
Figura 9 – PLN de periódicos eletrônicos	62
Figura 10 – Estrutura Geral Rede Cariniana	64
Figura 11 – Parceiros integrais da Rede Cariniana	68
Figura 12 – Processo de ingestão	73
Figura 13 – Resultado do processo de Poll	75
Figura 14 – Interface administrativa da caixa LOCKSS UFRN	76
Figura 15 – Configuração de periódico	77
Figura 16 – Controle de acesso administrador	77
Figura 17 – Informação de <i>proxy</i>	78
Figura 18 – <i>Status</i> do <i>daemon</i>	79
Figura 19 – Painel de <i>debug</i>	79
Figura 20 – Opções de acesso de conteúdo	81
Figura 21 – visualização do processo <i>crawl</i>	96
Figura 22 – Visualização – Origem da UA	96
Tabela 1 – Elementos de extensão para Dublin Core – RI UFRN	60
Tabela 2 – Configurações de Servidor	70
Tabela 3 – Elementos para descrição das Teses e Dissertações	90
Tabela 4 – Elementos <i>Dublin Core</i> obrigatórios	91
Tabela 5 – Conjunto de requisitos	101
Tabela 6 – Boas práticas	102

Lista de abreviaturas e siglas

AIP – Archival Information Package
BCZM – Biblioteca Central Zila Mamede
BDTD – Biblioteca Digital de Teses e Dissertações
BOAI – Budapest Open Access Initiative
CONSEPE – Conselho Superior de Ensino
DIP – Dissemination Information Package
FINEP – Financiadora de Estudos e Projetos
HNRS – Handle.Net Registry Service
IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia
JBB – Jardim Botânico de Brasília
JPEG – Joint Photographic Experts Group
JSPUI – Java Server Pages User Interface
LOCKSS – Lots Of Copies Keep Stuff Safe
MCTI – Ministério da Ciência, Tecnologia e Inovação
METS – Metadata Encoding & Transmission Standard
MIT – Massachusetts Institute of Technology
MTD-BR – Padrão Brasileiro de Metadados para Teses e Dissertações
OAI – Open Archive Initiative
OAI-PMH – Open Archives Initiative Protocol for Metadata Harvesting
OCLC – Online Computer Library Center
OJS – Open Journal Systems
P2P – Peer-to-peer
PDD – Preservação Digital Distribuída
PDF – Portable Document Format.
PLN – Private LOCKSS Network
PREMIS – Preservation Metadata: Implementation Strategies
RD – Repositório Digital
RI UFRN – Repositório Institucional da UFRN
RLG – A Research Libraries Group
ROAR – Registry of Open Access Repositories
SEER – Sistema Eletrônico de Editoração de Revistas

SINFO – Superintendência de Informática
SIP – Submission Information Package
SPARC – Scholarly Publishing and Academic Resources Coalition
SPD – Sistemas de Preservação Digital
SRB – Storage Resource Broker
SRD – Setor de Repositórios Digitais
TEDE – Sistema de Publicação Eletrônica de Teses e Dissertações
TIC – Tecnologias de Informação e Comunicação
UA – Unidade de Arquivamento
UEMA – Universidade Estadual do Maranhão
UFPB – Universidade Federal da Paraíba
UFRN – Universidade Federal do Rio Grande do Norte
UNB – Universidade de Brasília
UNICAMP – Universidade de Campinas
USP – Universidade de São Paulo
XML – eXtensible Markup Language

Sumário

Resumo	4
Lista de Figuras e Tabelas	6
Lista de abreviaturas e siglas	7
Introdução	13
1. Enquadramento e motivação	13
2. Objetivos e resultados esperados.....	16
3. Abordagem metodológica	16
4. Estrutura da dissertação	18
1. Preservação digital de longo prazo	20
1.1. Preservação digital.....	20
1.2. Objeto digital	23
1.3. Autenticidade e integridade	24
1.4. Estratégias de preservação	25
1.4.1. Estratégias estruturais	26
1.4.1.1. Modelo de referência OAIS	27
1.4.1.2. Metadados.....	30
1.4.2. Estratégias operacionais	33
1.4.2.1. Preservação de tecnologia.....	33
1.4.2.2. Emulação	34
1.4.2.3. Encapsulamento	34
1.4.2.4. Migração	35
1.4.2.5. Pedra de Roseta	38
1.5. Preservação digital distribuída.....	38
1.5.1 LOCKSS	40
2. Repositórios digitais de acesso livre	42
2.1. Acesso livre	42
2.2. Repositórios digitais	43
2.3. Repositórios institucionais	46
2.4. Tecnologias e ferramentas.....	47
2.4.1. DSpace.....	48
2.5. Síntese crítica	51

3. Diagnóstico e análise do contexto institucional de cooperação	52
3.1. Repositório Institucional da Universidade Federal do Rio Grande do Norte....	52
3.1.1. Arquitetura de informação	55
3.1.2. Formatos e identificação de arquivos	56
3.1.3. Depósito e Licenças.....	57
3.1.4. Fluxo de submissão.....	58
3.1.5. Metadados	59
3.2. Rede Cariniana.....	60
3.3. Âmbito da cooperação	66
3.3.1 Tipos de parcerias	66
3.3.2. Componentes Tecnológicos.....	69
3.3.3. Processos de ingestão, armazenamento, gestão e acesso aos conteúdos	71
3.3.3.1. Ingestão	72
3.3.3.2. Armazenamento	74
3.3.3.3. Gestão	75
3.3.3.4. Acesso	80
3.3.4. Testes de integração	82
4. Conjunto de Requisitos	84
4.1. Requisitos	84
4.1.1. Escopo.....	86
4.1.2. Metadados.....	88
4.1.2.1. Padronização dos Formulários	88
4.1.2.2. Elementos obrigatórios	90
4.1.3. Formato.....	92
4.1.4. Nomeação de arquivos.....	94
4.1.5. Arquitetura caixa LOCKSS.....	95
4.2. Boas Práticas	97
4.2.1. Licença de acesso e uso.....	98
4.2.2. Monitoramento da caixa LOCKSS	98
4.2.3. Verificação de vírus	99
4.3. Quadro Síntese.....	99
Considerações Finais.....	102
Referências Bibliográficas	109

ANEXOS	117
APÊNDICES	123

Introdução

1. Enquadramento e motivação

Ao longo dos séculos vem sendo possível preservar, de forma satisfatória, informação armazenada em suportes físicos, como livros, periódicos e outros documentos impressos, permitindo a salvaguarda da memória e o acesso à informação contida nestes materiais. Para sua preservação e conservação foi necessário pois, aprender a lidar com os perigos e ameaças ao suporte como agentes físicos, químicos ou biológicos, adotando estratégias voltadas ao objeto em si, uma vez que suporte e conteúdo representam uma única identidade indissociável.

O advento e evolução das Tecnologias da Informação e Comunicação - TIC proporcionou o surgimento e conseqüentemente um grande crescimento na quantidade de informação disponível eletronicamente seja ela nascida nesse formato ou migrada para meio digital em diferentes suportes. Contrariamente ao que foi dito sobre a preservação dos documentos físicos tradicionais, ainda não se pode afirmar que os documentos em meio digital estão sendo preservados satisfatoriamente de forma a garantir não só a salvaguarda da memória, como também o acesso de longo prazo a esses materiais, sem perdas no que diz respeito a sua autenticidade e integridade.

Quando se trata da informação em meio digital, a preservação deixa de ter o foco voltado para o objeto físico em si, e toma uma dimensão mais ampla, voltada para todo o ambiente tecnológico no qual se insere. Nesse contexto informação e suporte podem estar dissociados, o que torna necessário o desenvolvimento de estratégias e ações voltadas para a salvaguarda dos conteúdos e também desse ambiente. Esse, por sua vez, está sujeito a uma fragilidade inerente às novas tecnologias, a obsolescência tecnológica, resultante das constantes buscas por inovação que objetivam tornar os suportes e formatos eletrônicos cada vez mais eficientes e eficazes.

Essa constante renovação torna as TIC instáveis. E são essas tecnologias que estão sendo utilizadas de forma massiva em nível mundial em todas as esferas da

sociedade, proporcionando alterações significativas nas formas de produzir, armazenar, preservar, aceder e comunicar a informação. Nessa conjuntura, documentos digitais podem ser perdidos com a mesma rapidez com que são criados.

No caso das instituições de ensino superior e de pesquisa, esse contexto torna-se ainda mais relevante, uma vez que essas instituições concentram pesquisadores de alto nível, sendo, portanto, consideradas as principais responsáveis pela produção de conhecimento científico, resultantes das atividades fundamentais de ensino e investigação, a nível mundial (Leite e Costa, 2006). O conhecimento gerado no âmbito dessas instituições vem sendo registrado, e em maior quantidade a cada ano, em formatos eletrônicos em meio digital.

Nessa condição, as atividades ligadas à disponibilização e salvaguarda desse conhecimento, anteriormente voltadas para os suportes tradicionais (como livros e revistas, por exemplo), sofrem forte impacto provocado pelas TIC, migrando para o ambiente tecnológico de forma massiva. Isso já é realidade nas universidades em todo o mundo, que direcionam sua produção para o ambiente digital, seja ele de acesso restrito ou livre, como é o caso dos Repositórios Digitais de Acesso Livre. Esses repositórios foram amplamente adotados como forma de processar, tornar acessível e também salvaguardar a produção científica das instituições, permitindo o acesso imediato e amplo ao conhecimento científico gerado nessas instituições.

Porém, como escrito anteriormente, esse ambiente tecnológico é frágil e evolui de forma rápida e imprevisível expondo os formatos e suportes à obsolescência. Assim, no contexto das universidades e dos repositórios digitais, e tendo em vista a importância da informação para o desenvolvimento das atividades consideradas fundamentais para essas organizações, e ainda, uma dependência cada vez maior das TIC, a questão da preservação digital de longo prazo ganha fundamental importância.

Nas Universidades públicas brasileiras os repositórios digitais de acesso livre surgem a partir dos anos 2000 por iniciativa do Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT¹, com a Biblioteca Digital Brasileira – BDB (Moreno; Leite; Márdero Arellano, 2006). E, a partir de 2004 o IBICT iniciou os trabalhos de adaptação do *software* DSpace para a criação de Repositórios Institucionais em universidades públicas brasileiras.

¹ Disponível em: <http://www.ibict.br/>

Na Universidade Federal do Rio Grande do Norte – UFRN² o primeiro repositório digital foi criado em 2006 com objetivo de gerir, armazenar e dar acesso as teses e dissertações defendidas na universidade, denominado Biblioteca Digital de Teses e Dissertações – BDTD. Posteriormente, foram criados o Portal de Periódicos Eletrônicos (2009) com o objetivo de gerir todas as revistas científicas produzidas no âmbito da UFRN como também todo o processo editorial das mesmas, o Repositório Institucional – RI (2010), para a produção científica dos membros da comunidade acadêmica, e mais recentemente, a Biblioteca Digital de Monografias – BDB (2015), para as monografias produzida no âmbito dos cursos de bacharelado, licenciatura, e Especializações em nível *Lato Sensu* da universidade.

Buscando estar em consonância com esse novo contexto e, reconhecendo a importância de garantir o acesso de longo prazo ao conhecimento científico gerado nessas instituições, no ano de 2013, o IBICT iniciou o processo de criação e implantação de uma rede de serviços de preservação digital de documentos eletrônicos brasileiros, a Rede Cariniana³, com o objetivo de garantir a salvaguarda e o acesso contínuo em longo prazo⁴ desse conhecimento.

Diante desse cenário e da importância estratégica de sua produção científica, e entendendo que um esforço individual para preservar os conteúdos armazenados em seus repositórios digitais seria dispendioso e moroso, a UFRN aderiu em 2014, por meio de acordo de cooperação, à Rede Cariniana. Esse acordo objetiva garantir a salvaguarda e o acesso de longo prazo a esses conteúdos. Nessa perspectiva, o presente estudo justifica-se pela clara necessidade da adequação do Repositório Institucional da UFRN – RI UFRN, para tornar possível sua adesão à rede e a incorporação de seus conteúdos digitais. Pretende-se pois, apoiar esse processo propondo um conjunto de requisitos mínimos voltados para a integração entre o RI UFRN e a Rede Cariniana, no âmbito do referido acordo de cooperação.

² Disponível em: <http://www.sistemas.ufrn.br/portal/PT/>

³ Disponível em: <http://cariniana.ibict.br/>

⁴ “longo prazo define o espaço de tempo determinado pelo acesso continuado aos recursos digitais ou pelo menos à informação neles contida, indefinidamente” (Saramago, 2004).

2. Objetivos e resultados esperados

Este estudo tem como principal objetivo desenvolver um conjunto de requisitos mínimos, direcionado para a UFRN e seu repositório, no âmbito da cooperação com a Rede Cariniana. Pretende-se, assim, apoiar o acordo de cooperação entre a UFRN e o IBICT que visa a preservação do acervo digital da universidade por meio da Rede Cariniana. Para atingir o principal objetivo, realizou-se um estudo exploratório acerca do RI UFRN e da Rede Cariniana, com a finalidade de identificar as particularidades de cada uma dessas estruturas, bem como as relações entre ambas. Com o intuito de atingir o objetivo principal, o estudo irá apoiar-se nos seguintes objetivos específicos:

- ✓ Explorar o RI UFRN;
- ✓ Explorar a Rede Cariniana;
- ✓ Reconhecer relações tecnológicas existentes entre o RI UFRN e a Rede Cariniana;
- ✓ Identificar os meios que podem viabilizar a incorporação do acervo digital armazenado no RI UFRN na Rede Cariniana;
- ✓ Identificar os meios que permitam assegurar a autenticidade e integridade do acervo;
- ✓ Identificar os processos utilizados pela Rede Cariniana, que irão possibilitar a salvaguarda e acesso de longo prazo do acervo digital da UFRN;
- ✓ Desenvolver um conjunto de requisitos para promover a integração entre o RI UFRN e a Rede Cariniana.

3. Abordagem metodológica

Seguindo a tipologia definida por Gerhardt e Silveira (2009), este trabalho se constitui em uma investigação aplicada, e exploratória, uma vez que objetiva proporcionar maior familiaridade com um problema. Quanto a sua abordagem, é

uma investigação de natureza qualitativa, pois não se preocupa com uma representatividade numérica, mas com o aprofundamento da compreensão de um contexto específico, o âmbito da cooperação RI UFRN – Rede Cariniana.

Desta forma, a pesquisa exploratória que corresponde à parte inicial deste trabalho objetivou maior compreensão acerca dos repositórios digitais e da preservação digital, identificando relações e principais conceitos. Posteriormente, com o intuito de alcançar o objetivo proposto o estudo se desenvolveu em quatro fases: (1) a coleta de dados relativos ao RI UFRN e a Rede Cariniana; (2) análise das informações coletadas; (3) primeiros testes de integração entre o DSpace e o LOCKSS; e (4) desenvolvimento do conjunto de requisitos.

Na primeira etapa de coleta os dados foram recolhidos por meio de leitura/análise de fontes informacionais relativas ao RI UFRN e a Rede Cariniana. Essa leitura/análise foi realizada de forma direcionada, ou seja, buscando identificar as estruturas de funcionamentos; os componentes tecnológicos; as políticas; os padrões; e os requisitos de preservação.

De acordo com González Rey (2005), é por meio do uso de instrumentos para recolha de dados que se legitima a produção de informação na pesquisa. Esses instrumentos podem ser individuais, relacionados com a coleta por meio de questionários ou entrevistas aplicados individualmente ou em grupos. Nesse sentido, para a segunda fase da coleta de dados, utilizou-se a entrevista semiestruturada, como instrumento individual. Esse tipo de entrevista parte da elaboração de um guião flexível, que pode ser adaptado as necessidades uma vez que não segue nenhum modelo rígido, possibilitando que questões que não foram pré-determinadas sejam acrescentadas de acordo com as necessidades do entrevistado ou do entrevistador, desde que julguem pertinentes (Martins, 2010).

Assim, guiões para realização de entrevistas semiestruturadas foram elaborados com base em lacunas deixadas na primeira etapa da coleta de dados. As entrevistas foram direcionadas para os Coordenadores do RI UFRN e da Rede Cariniana, Analista de sistemas da BCZM / UFRN e ao Técnico de TI da Rede Cariniana.

Conforme Cação (2009) uma análise qualitativa de informações visa obter, por meio de procedimentos sistemáticos, dados que permitam inferir sobre o tema

tratado. Dessa forma, a segunda fase se constituiu na análise qualitativa das informações coletadas. Essa análise possibilitou compreender as estruturas de forma sistêmica e identificar requisitos, componentes, padrões adotados e relações, no RI UFRN e na Rede Cariniana. A pesquisa exploratória, juntamente com a compreensão das referidas estruturas, forneceram o embasamento necessário à criação do conjunto de requisitos, principal objetivo deste estudo.

Na terceira fase, foram realizados os primeiros testes de integração entre o DSpace e o LOCKSS. Os testes foram coordenados pelo especialista da *Stanford University*, Thibault Yoann Guicherd-Callin e realizados nas dependências do IBICT, em Brasília, entre os dias 21 e 25 de Março de 2015⁵. Para esses testes, foram utilizados os bancos de dados do RI UFRN e do Repositório Institucional da Universidade de Brasília – UNB.

A quarta fase compreendeu a criação de um conjunto de requisitos mínimos para a integração entre o RI UFRN e a Rede Cariniana, a ser proposto à UFRN.

4. Estrutura da dissertação

O presente documento encontra-se estruturado em seis partes. A primeira parte introduz o tema, apresenta o contexto e a motivação, como também, os objetivos e procedimentos metodológicos adotados.

A segunda e terceira partes correspondem respectivamente ao primeiro e segundo capítulos, que tratam da revisão da literatura. Aborda a preservação digital de longo prazo, origem, importância, conceito, estratégias e abordagens de preservação digital. E também, conceitos que permeiam a preservação digital tais como objeto digital, autenticidade e integridade. Por fim, discute a preservação digital distribuída e o *software* LOCKSS. Bem ainda, aborda questões relativas aos repositórios digitais de acesso livre, contexto no qual surgiram, conceitos, tipologia e importância desses repositórios

⁵ Durante esse período foram realizadas diversas apresentações sobre o LOCKSS e a Rede Cariniana, além dos testes de integração.

A quarta parte, capítulo três, expõe e analisa o contexto organizacional de cooperação. Caracteriza o RI UFRN e a Rede Cariniana, com seus devidos processos e serviços, especifica os contextos tecnológicos de ambas e as relações existentes. O capítulo apresenta ainda informações relativas aos primeiros testes de integração entre o DSpace e o LOCKSS.

A quinta parte, corresponde ao capítulo quatro, é dedicada ao desenvolvimento do conjunto de requisitos, objetivo principal deste estudo.

Por fim, são apresentadas as considerações finais sobre o estudo, faz-se uma verificação relativa ao objetivo proposto com o intuito de analisar seu cumprimento. E, de seguida, são apresentadas as referências bibliográficas e anexos.

1. Preservação digital de longo prazo

Este capítulo aborda questões pertinentes a preservação digital de longo prazo, tais como origem, importância, conceito, estratégias e abordagens de preservação digital. Aborda ainda alguns dos principais conceitos que permeiam a preservação digital, tais como objeto digital, autenticidade e integridade. E, por fim, discute a preservação digital distribuída e o *software* LOCKSS.

1.1. Preservação digital

Em 1997 Terry Kuny publicou um artigo intitulado “*A Digital Dark Ages?*”. Ele, fez um comparativo entre o atual contexto, que chamou de “*hyperbolic environment of technology*” e a época medieval, onde monges desempenharam papel fundamental para a preservação e distribuição de livros. O autor se apropriou da comparação para afirmar que a sociedade estava vivendo “no meio da idade das trevas digital”, para enfatizar a importância dos profissionais da informação em relação a esse novo contexto de informação eletrônica e de grande volumes de dados, e também da necessidade de que estes profissionais, assim como os monges, ajam no sentido de preservar essa informação. Muito embora, segundo ele, parte dessa informação, codificada e escrita por via eletrônica, será perdida para sempre (Kuny, 1997).

Segundo a IBM, todos os dias são criados cerca de 2,5 quintiliões de bytes de dados, e 90% dos dados no mundo hoje, foram criados nos últimos dois anos⁶. Somente o Google, processa cerca de 20 peta bytes de dados por dia, através de mais de um milhão de servidores espalhados pelo mundo, monitorando cerca de 7,2 bilhões de páginas web (Sagiroglu e Sinanc, 2013). A *EMC Corporation*⁷, como resultado de um estudo anual que quantifica e prevê o volume de dados produzidos mundialmente, aponta que o universo digital vem dobrando de tamanho a cada dois

⁶ Disponível em: <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>

⁷ Disponível em: <http://portugal.emc.com/leadership/index.htm>

anos, e que se “multiplicará por dez entre 2013 e 2020 – de 4,4 triliões de gigabytes para 44 triliões de gigabytes” (EMC, 2004). Esses dados correspondem a troca de mensagens, envio de fotos, arquivos, emails, áudios, vídeos, interações em redes sociais, registros de transações comerciais, sinais telefônicos e de GPS – *Global Positioning System*, dados científicos, para citar apenas alguns, gerados nos mais diversificados setores da sociedade.

Este quadro expõe não só a importância das TIC e a consequente dependência dessas tecnologias, como também, a da informação eletrônica para a atual sociedade, enquanto recurso vital. Sayão enfatiza que a “informação é um insumo essencial para o desenvolvimento da ciência e, consequentemente, para o desenvolvimento social e econômico de um país” (Sayão e Marcondes, 2009:19). Mas, o insumo essencial só terá significado se puder ser recuperado.

Ocorre que, atualmente, esse insumo encontra-se mediado por *hardware* e *software*, sendo, portanto, sua criação e acesso dependentes de um meio ambiente tecnológico que por natureza, é transitório. Essa característica se deve ao fato das tecnologias digitais sofrerem constante evolução, estando por esta razão sujeitas a obsolescência tecnológica. Hedstrom (2001) observa que, a cada nova geração de *hardware* e *software* a geração anterior tende a ser obsoleta. Uma vez registrada em um dado formato a informação obrigatoriamente respeita as regras deste formato, o que permitirá seu acesso e uso. Porém, à medida que o *software* evolui esses formatos sofrem alterações (Ferreira, 2006). Isso significa que informação transferida de gerações anteriores para geração mais atual, podem ser corrompidas (Hedstrom, 2001), o que as torna, portanto, vulneráveis.

Esses dois fatores descritos, a dependência e a obsolescência tecnológicas são apontados por Rodrigues como os mais preocupantes da ‘era tecnológica’ (Rodrigues, 2004). Nessa nova ‘era’ é preciso garantir não só a longevidade como também a integridade dessa informação, o que torna “premente a implementação de técnicas e de políticas concertadas que vão no sentido de garantir a perenidade e a acessibilidade a este tipo de informação” (Ferreira, 2006:20). Essas técnicas e políticas são compreendidas no âmbito da Preservação Digital de Longo Prazo.

A preservação digital surge, portanto, como um conjunto de processos que visa “garantir que a informação digital permaneça acessível e com qualidades de

autenticidade suficientes para que possa ser interpretada no futuro recorrendo a uma plataforma tecnológica diferente da utilizada no momento da sua criação”, e também que a comunicação entre emissor e receptor seja possível através do tempo e do espaço (Ferreira, 2006:20). De acordo com o *Guidelines for the preservation of digital Heritage*⁸ a preservação digital deve ser utilizada para descrever processos envolvidos na manutenção de informação e demais tipos de património existentes em um formato digital (UNESCO, 2003).

Segundo o *InterPARES Project*⁹ a preservação digital é um

“Processo específico de manutenção de materiais digitais ao longo do tempo e através de diferentes gerações de tecnologia, independentemente do local de armazenamento; Conjunto de ações gerenciais e técnicas exigidas para superar as mudanças tecnológicas e a fragilidade dos suportes, garantindo o acesso e a interpretação de documentos digitais pelo tempo que for necessário”(InterPARES, http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=p&term=276).

Para a *Digital Preservation Coalition* (2008), “refere-se a todas as ações necessárias para manter o acesso aos materiais digitais para além dos limites de falha de média ou a mudança tecnológica”. Diante dessas definições, podemos colocar como principal objetivo da preservação garantir que o património digital seja salvaguardado de modo a permanecer acessível por longo prazo e de forma autêntica. Para isso, faz-se necessário a aplicação de políticas, estratégias e ações, que vão ao encontro deste objetivo.

Porém, antes de adentrar nessas especificações, se faz necessário compreender o que deve ser preservado. Isso permitirá uma melhor compreensão e enquadramento das técnicas e padrões que possibilitarão a preservação digital de longo prazo e ainda, compreender como, e se é possível, garantir que esta informação esteja acessível, íntegra e autêntica.

⁸ Disponível em: <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>

⁹ International Research on Permanent Authentic Records in Electronic Systems – InterPARES, “é uma ação colaborativa internacional de pesquisa sobre documentos arquivísticos autênticos em sistemas informatizados” Disponível em: <http://www.interpares.org/welcome.cfm>

1.2. Objeto digital

Objeto digital de acordo com Ferreira “pode ser definido como todo e qualquer objeto de informação que possa ser representado através de uma sequência de dígitos binários” (Ferreira, 2006: 21). A definição apontada pelo autor é bastante alargada e aplica-se tanto a informação nado-digital¹⁰, quanto à informação digitalizada¹¹ e de diferentes tipos como imagens digitais, documentos de texto, bases de dados, entre outros. Logo, essa definição se inter-relaciona com as definições descritas anteriormente para o termo documento digital.

Tendo em conta o contexto dos arquivos e bibliotecas digitais Marcondes define esses mesmos objetos como “um registro de informação codificado digitalmente, consistindo de conteúdo informacional, metadados e identificador” (Marcondes, 2005:18). Essa definição corrobora com a apontada pelo *Dictionary for Preservation Metadata – PREMIS*, que assume o objeto digital como uma espécie de ‘pacote’ formado pela combinação de identificador, metadados, e a informação em si, o conteúdo informacional (PREMIS Editorial Committee, 2008).

Para Thibodeau (2002), os objetos digitais devem ser vistos como uma entidade múltipla, ou seja, devem ser compreendido em três classes. Dessa fora, cada objeto digital é um objeto físico, um lógico, e um objeto conceitual, e cada uma dessas classes pode possuir propriedades significativamente diferentes. O objeto físico corresponde a uma inscrição de sinais sobre um dado suporte, ou seja, um meio físico onde as codificações lógicas dos bits são registradas. O objeto lógico corresponde um conjunto de informação organizada em estruturas de dados, reconhecida e processada por alguma aplicação de *software*. E o conceitual é o objeto que adquire significado para o ser humano, é como ele é reconhecido e compreendido pelo utilizador. (Pinto, 2009; Thibodeau, 2002; *InterPARES Project Terminology Database*).

O *Guidelines for the preservation of digital heritage*, propõe abordar o objeto de quatro formas: como fenómenos físicos, codificações lógicas, objetos conceituais

¹⁰ “informação nascida num contexto tecnológico digital” (Ferreira, 2006: 21)

¹¹ “informação digital obtida a partir de suportes analógicos” (Ferreira, 2006: 21)

compreensíveis para o ser humano e como um conjunto de elementos essenciais que devem ser preservados, permitindo que seja oferecido aos utilizadores o essencial do objeto (National Library of Australia, 2003).

Diante disso, entendemos aqui o documento digital ou eletrônico como sendo o conteúdo informacional, criado ou migrado para o meio digital, e que portanto, necessita de um contexto tecnológico, composto por *hardware* e *software*, para se tornar acessível. E assumimos o objeto digital, dada a complexidade e pluridimensionalidade, como um ‘pacote’, um conjunto de dados formado por identificador, metadados, e conteúdo informacional, que juntos constituem uma unidade de sentido.

1.3. Autenticidade e integridade

Hofman (2002) refere que o conceito de autenticidade pode assumir diferentes significados que vão sendo atribuídos de acordo com o domínio que o manipula. Por esta razão Ferreira (2009) afirma não haver um consenso para o conceito. Em linhas gerais a autenticidade pode ser descrita como a capacidade de um objeto digital ser aquilo a que se propõe (Digital Preservation Coalition, 2008; Ferreira, 2009; MacNeil, 2002). Na Terminology Database, é referida como “credibilidade de um documento arquivístico enquanto tal, isto é, a qualidade de um documento ser o que diz ser e de estar livre de adulteração e corrupção” (InterPARES. Terminology Database, 2014).

Quanto a integridade, o Glossário da Sociedade da Informação define como sendo uma “garantia de que os dados ou a informação não sejam alterados de modo não autorizado” (Associação Para Promoção e Desenvolvimento da Sociedade da Informação, 2011:69). Afirmar que um objeto digital é íntegro, significa dizer que os dados que o compõe não foram corrompidos ao longo do tempo, ou em trânsito, ou seja, o que é acedido representa o mesmo conjunto de sequência de bits que existiam quando da sua criação (Lynch, 2000).

No âmbito do projeto InterPARES a integridade é vista como a qualidade de ser completo e inalterado em todos os aspectos essenciais (InterPARES Terminology Database, 2014). Esses aspectos essenciais, ou propriedades significativas são “características que um objeto digital deve manter ao longo do tempo, independentemente das ações e estratégias de preservação levadas a cabo” (Ferreira, 2011:36). Quanto mais propriedades significativas forem atribuídas aos objetos digitais, maior deverá ser a infraestrutura necessária para suportar a sua preservação (Ferreira, 2009).

Em teoria, para cada objeto digital um conjunto de propriedades significativas deveria ser atribuído, porém, como indica Ferreira (2009), embora desejável essa atribuição não é economicamente viável, sendo necessária a criação de políticas de preservação que indiquem para cada classe de objetos um conjunto de propriedades que garantam a integridade dos mesmos. No entanto, a verificação dessa integridade consiste em comparar o objeto mais recente, com o que foi originalmente criado. Nesse sentido, testar a integridade pode ser visto como um subproduto de avaliações de autenticidade (Lynch, 2000).

Em um ambiente digital, e em um quadro ideal, questões como essas devem ser analisadas de antemão uma vez que podem interferir ou influenciar a funcionalidade dos aplicativos de *software* escolhidos para a preservação digital (Hofman, 2002). Objetos digitais requerem, portanto, políticas, estratégias e métodos de preservação com base em modelos e padrões visando permitir que se mantenham autênticos e íntegros para que possam assim ser acedidos.

1.4. Estratégias de preservação

A preservação digital supõe a adoção de um conjunto coerente de tecnologias de informação, instalações, pessoas, sistemas, processos e procedimentos que atuem em todas as dimensões do objeto digital, buscando: colaborar com os criadores da informação no que se refere a aplicação de normas que prolonguem a vida dos meios de acesso e na redução de problemas; identificar o que deve ser preservado; manter a

informação em lugar seguro e estável; fazer uso de metadado estruturado; proteger os dados quanto a sua integridade e autenticidade; adotar meios apropriados para permitir o acesso permanente; e, gerir os programas de preservação para que cumpram que seus objetivos de forma econômica, oportuna, global, dinâmica e responsável (National Library of Australia, 2003).

Dessa forma, segundo Thibodeau (2002), um sistema ideal de preservação seria um canal de comunicação neutro para a transmissão de informações para o futuro, e este canal não deve alterar ou corromper a mensagem transmitida. Para tal, faz-se necessário a aplicação de estratégias de preservação digital que visem garantir essa transmissão, como também a integridade e autenticidade da informação. Contudo, essas estratégias estão condicionadas aos conhecimentos que temos hoje sobre as tecnologias, e à obsolescência tecnológica, o que faz com que todas as estratégias de preservação existentes sejam questionadas (Márdero Arellano, 2009). Para Rosenthal et al (2005) o sucesso de um sistema de preservação digital permanecerá para sempre não comprovado, já o fracasso, se tornará evidente em tempo finito.

Não existe uma categorização consensual acerca dessas estratégias, no entanto, de acordo com Mardero Arellano (2004), podem ser agrupadas em dois tipos: estruturais e operacionais. As estruturais dizem respeito aos investimentos iniciais para implementação de algum processo de preservação, como adoção de padrões, elaboração de normas, metadados de preservação, infraestrutura. As operacionais são as medidas concretas aplicadas aos objetos digitais, como preservação de tecnologia, emulação, encapsulamento e migração.

1.4.1. Estratégias estruturais

As ações estruturais precedem as medidas concretas para preservação digital de longo prazo. Faz-se necessária a adoção de políticas de preservação e padrões ou modelos que possam nortear o processo de preservação. As políticas devem fornecer orientações legais como também acerca do funcionamento do processo de

preservação como um todo, ficando a cargo de cada instituição elaborá-las com base em resoluções, normas, leis modelos e padrões. Relativamente aos padrões, destaca-se o modelo de referência OAIS e os metadados de preservação.

1.4.1.1. Modelo de referência OAIS

No contexto da preservação, o modelo de referência OAIS¹² - *Open Archival Information System* foi amplamente adotado como um padrão de fato em preservação digital. O surgimento do modelo se deu com a formação de um grupo de trabalho patrocinado pela OCLC - Online Computer Library Center ¹³ e pelo RLG - Research Libraries Group, com o intuito explorar a formação de consensos na área de metadado de preservação e desenvolver de forma colaborativa uma estrutura de metadados que pudesse ser aplicada a uma ampla gama de atividades relativas a preservação digital (OCLC, 2002). Em Fevereiro de 2003 o modelo foi aprovado como norma internacional ISO - International Organization for Standardization - ISO 14721: 2003¹⁴.

O OAIS é um modelo de referência que fornece uma terminologia própria, facilitando a comunicação entre os envolvidos na preservação, e proporcionando condições para o estabelecimento de projeto de preservação para qualquer tipo de documento, digital ou analógico, e, por essa razão, não deve ser considerado como uma implementação específica (Márdero Arellano, 2008). O OAIS fornece um modelo conceitual (Figura 1) que identifica o funcionamento de sistemas de preservação digital como um todo. Nele, são identificadas quatro entidades externas: o produtor, que fornece ao sistema a informação a ser preservada; o administrador, que define políticas de administração e coordena o funcionamento do sistema; o consumidor, que acede e utiliza as informações arquivadas; e o arquivo propriamente dito, o pacote de informação. E ainda, seis entidades funcionais: ingestão;

¹² Traduzida para o português como SAAI - Sistema Aberto de Arquivamento de Informação

¹³ Disponível em: <https://oclc.org/home.en.html>

¹⁴ Esta norma foi revista em 2012. Disponível em: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57284

repositórios de dados; gestão de dados; acesso; planejamento de preservação; e administração (Márdero Arellano, 2008; OCLC, 2002;).

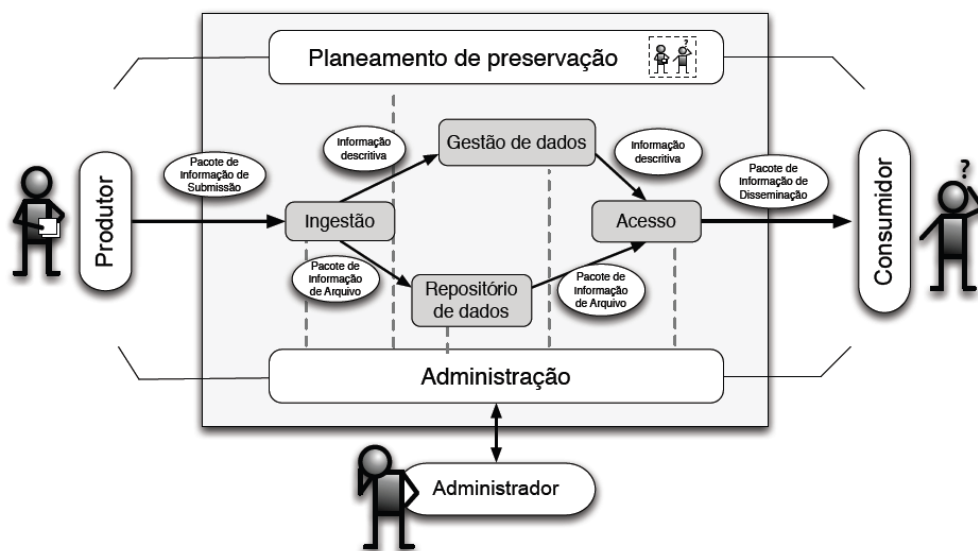


Figura 1 – Modelo de referência OAIS (Ferreira, 2009:19)

A ingestão é responsável por receber o *Submission Information Package* – SIP. Ela funciona como intermediário, recebendo o SIP, gerando a partir deste o *Archival Information Package* – AIP e o envia para o Repositório de dados. Nesse processo é extraída a informação descritiva necessária para a pesquisa, que posteriormente é enviada à Gestão de Dados (Ferreira 2009; Ferreira, 2011). A entidade Repositório de Dados, é responsável por serviços e funções que permitem o armazenamento, conservação e recuperação do AIP. A Gestão de Dados fornece serviços e funções necessárias para criar e manter bases de dados de metadados descritivos, utilizados para identificar a informação armazenada. À entidade Acesso cabe os serviços e funções que dão suporte aos utilizadores, para que esses solicitem, localizem e acessem à informação armazenada. Essa informação é entregue ao utilizador sob a forma de *Dissemination Information Package* – DIP. Cabem ao Planejamento de Preservação os serviços e funções relativos ao monitoramento do ambiente OAIS, tais como: desenvolvimento e aplicação de estratégias de preservação; análise do ambiente externo que possam causar algum tipo de mudança que afete a preservação; Atualização de políticas; e elaboração de relatórios. Por fim, à entidade Administração cabe gerir o modelo como um todo, estabelecendo e

mantendo políticas e normas, fornecendo suporte as demais entidades. (Pacheco et al, 2008; Ferreira, 2009; Ferreira, 2011).

Como parte deste modelo conceitual, o OAIS fornece ainda, um modelo de informação (Figura 2) que considera dois tipos de objetos, físicos e digitais, que são referidos como objetos de dados. Ele descreve detalhadamente os requisitos de metadados associados à retenção de um objeto digital a longo prazo.

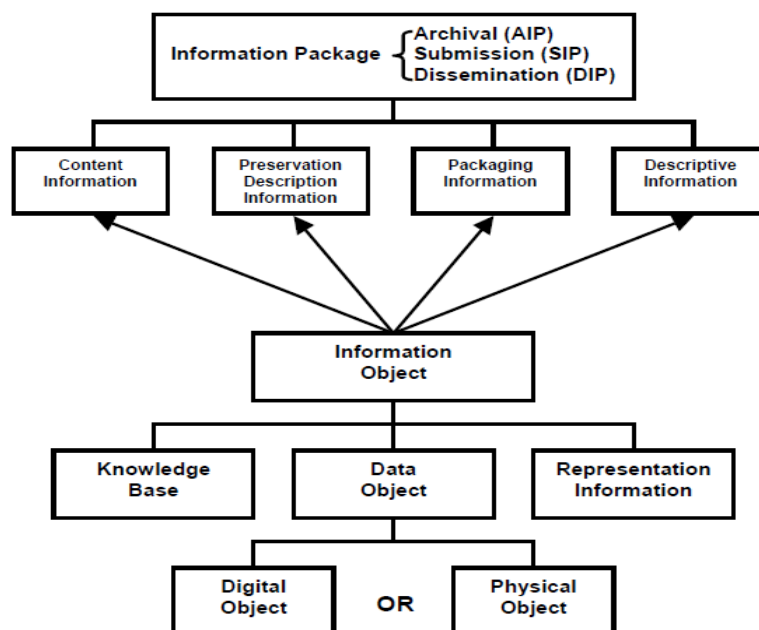


Figura 2 – Modelo de informação OAIS (OCLC/RLG Working Group on Preservation Metadata, 2001:11)

Como um modelo de referência, o OAIS permite que cada entidade que o adote tenha flexibilidade para determinar normas específicas de acordo com necessidades e particularidades suas e da comunidade na qual se insere, como por exemplo, normas para os pacotes de informação, gerenciamento de arquivo ou mesmo gerenciamento de dados (Hedstrom, 2001). O OAI é uma estrutura conceitual e propõe um modelo de informação para descrever, de forma ampla, metadados associados a retenção de objetos digitais. Dessa forma, ele não determina um conjunto de metadados para preservação digital de longo prazo, e sim um “modelo em que os padrões de metadados devem ser propostos” (Márdero Arellano, 2008:95). O modelo se coloca,

portanto, como um ponto de partida para que esquemas de metadados de preservação sejam criados. E, dessa forma, as instituições têm autonomia para criarem seus próprios padrões ou adotarem padrões já disponíveis.

1.4.1.2. Metadados

Metadados são comumente referidos como sendo dados sobre dados, ou ainda, informação sobre informação. A *National Information Standards Organization* (2004:1) apresenta uma definição mais detalhada onde conceitua metadados como informação estruturada que descreve, explica, localiza, ou de outra forma torna mais fácil a recuperação, uso ou gerenciamento de um recurso de informação. Trazendo o conceito para o ambiente da preservação digital, Saramago (2004), define metadado para preservação de longo prazo “como informação de apoio aos processos associados com a preservação digital de longo prazo”. Essa informação de apoio representa, por exemplo, informação técnica, administrativa, sobre estratégias de preservação, direitos de autor, dentre outras, fornecendo um alto nível de detalhamento dos objetos digitais, necessário para a salvaguarda, preservação, acesso e uso desses objetos.

Dessa forma, os metadados de preservação assumem papel fundamental na preservação digital de longo prazo. O OCLC/RLG Working Group on Preservation Metadata (2001), enfatiza que objetos digitais são mutáveis, e que, portanto, a garantia de autenticidade e integridade requer a manutenção, ao longo do tempo, de um histórico de alterações do objeto. Os metadados podem descrever esse histórico como também toda a história custodial dos objetos, ou seja, todo o seu percurso desde sua criação até sua incorporação em um repositório digital, e ainda descrever informação relativa ao ambiente tecnológico, necessária à correta apresentação e execução dos objetos (Ferreira, 2009). Assim, metadados de preservação tornam-se complemento necessário, sendo transversais a todas as estratégias operacionais. Além da preservação, desempenham papel igualmente importante para o intercâmbio de informação de preservação e para a interoperabilidade entre repositórios. (Sayão e Marcondes, 2009:39).

Três categorias¹⁵ são comumente atribuídas aos metadados, descritiva, administrativa e estrutural. Elas fornecem uma visão geral dos tipos de metadados que devem ser associados a um objeto digital. Metadados descritivos descrevem dados que visam a descoberta e identificação podendo incluir elementos como título, autor, resumo, palavras-chave e identificador persistente. Metadados estruturais apresentam informações que indicam como os objetos compostos por vários elementos são agrupados. E por fim, metadados administrativos, que fornecem informações para apoiar a gestão do ciclo de vida¹⁶ da informação, como quando e como ele foi criado, tipo de arquivo e outras informações técnicas do recurso (*National Information Standards Organization*, 2004; OCLC/RLG Working Group on Preservation Metadata, 2001).

O modelo de referência OAIS, forneceu as bases para criação de esquemas de metadados para dar suporte a preservação digital. A partir dele, diferentes conjuntos de metadados, com graus diferenciados de especificidades, têm surgido no intuito de abranger os diversos tipos de objetos digitais disponíveis e os distintos domínios de aplicação. Porém, cada iniciativa tende a desenvolver esquemas voltados para suas necessidades, limitando por vezes a aplicação dos esquemas ao contexto no qual foi criado. Dos esquemas de metadados desenvolvidos (ou em desenvolvimento) por variadas iniciativas, citamos aqui os mais referidos na literatura¹⁷ divididos em esquemas de aplicabilidade restrita a tipo de objeto ou domínio, e esquemas de ampla aplicabilidade, ou seja, que pode ser aplicado a variados tipos de objetos.

ESQUEMAS DE APLICABILIDADE RESTRITA A TIPO DE OBJETO OU DOMÍNIO:

- ✓ CDWA – Categories for the Description of Works of Art ¹⁸. Desenvolvido pela Art Information Task Force (AITF) para a descrição de objetos e imagens de arte;

¹⁵ No entanto, não há consenso acerca dessa categorização, porém, Sayão (2010) afirma que uma fração significativa de autores da área a utiliza.

¹⁶ O ciclo de vida da informação compreende a criação, seleção, identificação persistente, descrição e acesso, armazenamento e preservação (Saramago, 2004)

¹⁷ De acordo com a bibliografia utilizada para este quadro teórico de referência.

¹⁸ Disponível em: http://www.getty.edu/research/publications/electronic_publications/cdwa/

- ✓ EAD – *Encoded Archival Description*¹⁹. Voltado para descrever coleções de materiais de arquivo. Atualmente é mantido pela *Library of Congress*;
- ✓ ETD-MS – *Electronic Thesis and Dissertation Metadata Standard*²⁰. Conjunto de metadados utilizados para descrever teses ou dissertações eletrônicas, desenvolvido pela Networked Digital Library of Theses and Dissertations (NDLTD);
- ✓ LOM – Learning Object Metadata²¹. Modelo de dados para descrição de objetos de aprendizagem, desenvolvido pelo *IEEE Learning Technology Standards Committee*;
- ✓ MPEG – Moving Picture Experts Group²². O MPEG desenvolveu um conjunto de padrões voltados para a representação codificada de áudio e vídeo digital, e dados relacionados.

ESQUEMAS DE AMPLA APLICABILIDADE:

- ✓ Dublin Core Metadata Element Set²³;
- ✓ LMER - Long-term preservation Metadata for Electronic Resources²⁴;
- ✓ MARC 21²⁵;
- ✓ METS – Metadata Encoding and Transmission Standards²⁶;
- ✓ MODS – Metadata Object Description Schema²⁷;
- ✓ PREMIS Data Dictionary²⁸.

¹⁹ Disponível em: <http://www.loc.gov/ead/>

²⁰ Disponível em: <http://www.ndltd.org/standards/metadata>

²¹ Disponível em: <http://ltsc.ieee.org/wg12/>

²² Disponível em: <http://mpeg.chiariglione.org/>

²³ Disponível em: <http://dublincore.org/>

²⁴ Disponível em: <http://d-nb.info/975115073/34>

²⁵ Disponível em: <http://www.loc.gov/marc/>

²⁶ Disponível em: <http://www.loc.gov/standards/mets/mets-home.html>

²⁷ Disponível em: <http://www.loc.gov/standards/mods/>

²⁸ Disponível em: <http://www.loc.gov/standards/premis/bibliography.html>

1.4.2. Estratégias operacionais

Das estratégias operacionais, tratamos aqui das abordagens preservação de tecnologia, emulação, encapsulamento, migração e a Pedra de Roseta, por entendermos ser estas as mais comumente difundidas, utilizadas e, conseqüentemente mais referidas na literatura.

1.4.2.1. Preservação de tecnologia

A preservação da tecnologia consiste em conservar e manter o contexto tecnológico, *hardware* e *software*, usado originalmente na concepção do objeto. Essa estratégia pressupõe a criação de museus de tecnologia responsáveis por manter equipamentos obsoletos a fim de processar os objetos no seu ambiente original (Lee et al., 2002). Contudo, Ferreira (2011) aponta essa estratégia como sendo inviável em longo prazo por questões técnicas, espaciais e financeiras. Para manter os equipamentos se faz necessário espaço físico e manutenção, e estes, fatalmente se tornarão obsoletos.

Não só os equipamentos tornam-se obsoletos, como também o contexto que os envolve, como documentação e suporte de fornecedor. Para além disso, “também limita a portabilidade do recurso, uma vez que irá depender de *hardware* armazenados em locais específicos” (Lee et al, 2002). Entretanto, o autor aponta que os defensores desta estratégia argumentam que esta seria a única maneira de experimentar o objeto no seu estado original, ou seja, de forma fidedigna, sem comprometer a funcionalidade, aparência e autenticidade e, preservando o objeto nas suas dimensões física, lógica e conceptual (Lee et al, 2002).

1.4.2.2. Emulação

No âmbito do InterPARES, essa técnica é descrita como “reprodução do comportamento e resultados de *software* ou sistemas obsoletos através do desenvolvimento de um novo *hardware* ou *software* para permitir a execução do antigo *software* ou sistemas de computadores do futuro” (InterPARES Terminology Database). Nessa estratégia, um *software*, denominado emulador é utilizado para recriar o ambiente que permite apresentar o objeto no futuro, sem ter a tecnologia que o suportava mas emulando-a.

Para Rothenberg (1999), pode não ser possível preservar todos os atributos do objeto digital, porém, pode ser possível recriar o comportamento do documento da forma mais exata possível. Lee et al (2002) reforça que esta abordagem permite preservar a aparência e funcionalidades dos objetos digitais, com significativo grau de fidedignidade, o que a aproxima da preservação da tecnologia anteriormente descrita (Lee et al, 2002). Contudo, a natureza indefinida da tecnologia e a complexidade para se criar especificações para um emulador são consideradas duas das principais desvantagens da emulação (Lee et al, 2002). Sobre essa estratégia, Thibodeau (2002), ressalta que, ao passar do tempo o próprio emulador se tornará obsoleto, tornando-se necessário uma conversão para uma nova plataforma, ou ainda o desenvolvimento de um novo emulador.

1.4.2.3. Encapsulamento²⁹

A estratégia é vista por vezes como transversal a outras estratégias, como por exemplo, a emulação. Encapsular consiste “em preservar, juntamente com o objeto digital, toda a informação necessária e suficiente para permitir o futuro desenvolvimento de conversores, visualizadores ou emuladores” (Ferreira, 2009). Essa informação deve ser descrita de maneira formal e detalhada para que possam ser interpretadas no futuro. Assim, o encapsulamento mantém os objetos digitais

²⁹ Também referido como encapsulação

inalterados, em *stand by*, até ao momento em que seja revelado algum tipo de interesse pelos mesmos. Para essa descrição, idealmente, deve ser utilizada uma linguagem normalizada e compreensível por diferentes tecnologias, como XML por exemplo (Ferreira, 2011). Para Ferreira, Baptista e Ramalho (2005), essa linguagem XML permite descrever qualquer tipo de formato digital, e essa descrição pode garantir o sucesso da estratégia.

Em contrapartida, não é considerado prático, nem necessário, o encapsulamento dos meios de acesso utilizados ou seja, *software* e *hardware*. O encapsulamento normalmente deve incluir os metadados que descrevem ou estabelecem vínculos com as ferramentas necessárias (National Library of Australia, 2003). Como alternativa, pode-se incluir uma descrição do programa ou o computador, ou ainda um link para o mesmo, de modo que poderá ser restaurado no futuro, se necessário.

1.4.2.4. Migração

Diferentemente das técnicas anteriormente descritas, a migração não objetiva conservar os formatos de origem dos objetos digitais, e sim, representar o conteúdo intelectual, podendo não preservar aparência, funcionalidade ou contexto. Trata-se portanto de uma estratégia voltada para a preservação no nível conceitual do objeto. Segundo Lee et al (2002), a migração preserva a integridade dos objetos digitais permitindo aos utilizadores recuperar, exibir, e fazer uso deles mediante a constante evolução das tecnologias. Essa tem sido a estratégia de preservação mais amplamente implementada em repositórios digitais (Hedstrom, 2002; Márdero Arellano, 2009).

A migração consiste assim, em um conjunto de ações organizadas que visam a transferência periódica dos objetos digitais a partir de uma dada configuração de *hardware* e *software* para outra, ou ainda de uma geração tecnológica para outra subsequente (*Task Force on Archiving of Digital Information, Commission on Preservation and Access & Research Libraries Group*, 1996:6). Para tanto, recorre a *software* denominado conversor, que é responsável por converter dados registrados

em formatos obsoletos em novos, compatíveis com tecnologias mais atuais (Ferreira, Baptista e Ramalho, 2005).

Cada migração implica o desenvolvimento de ações personalizadas que envolvem a análise de formatos originais de arquivos, determinação de formato de destino e uso de conversor específico para cada uma. Isso faz com que o desenvolvimento desse conversor seja necessário, uma vez que conversores comerciais podem não atender as especificidades de cada formato por possuir configurações genéricas (Hedstrom, 2001).

Essa questão é apontada por Margaret Hedstrom como um dos principais motivos de crítica à migração. A autora diz que essa técnica não representa um processo único para todos formatos, ao contrário, requer abordagem individualizada, o que pode fazer com que o processo se torne lento. Um segundo motivo de crítica se deve ao fato da migração envolver algum tipo de transformação no fluxo de bytes original. Esse fluxo pode ser corrompido por erros de *software*, manuseio incorreto de dados ou falha mecânica dos dispositivos de entrada ou de saída. Essa transformação pode acarretar na perda de informação e funcionalidade, introdução de erros nos arquivos de destino, ou ainda algum tipo de mudança no processamento da informação para o utilizador (Hedstrom, 2001).

No entanto, para Ferreira, Baptista e Ramalho (2005), a principal vantagem da migração está em permitir que o utilizador comum consiga interpretar a informação sem que seja necessário o uso de outros mecanismos ou artefactos, para além de um computador pessoal. Hedstrom (2001) indica a existência de diferentes tipos de migração que afetam os objetos digitais de diferentes formas, tais como: migração a pedido; migração distribuída; atualização de versão; conversão para formatos concorrentes; e, normalização.

Na migração a pedido “a versão original do material é retida, e quando necessário, são aplicadas ferramentas próprias para converter o original para formatos requeridos pelo usuário” (Márdero Arellano, 2008:65). Nesse sentido, a migração irá ocorrer somente no momento em que a informação for solicitada. Nessa estratégia a informação é armazenada em nível de byte, para posterior conversão, quando for necessário. Nessa estratégia, a conversão deve sempre ser aplicada ao

objeto original, visando minimizar possíveis diferenças e perda de propriedades significativas (Ferreira, 2011; Ferreira, 2009).

A migração distribuída consiste na introdução de arquiteturas distribuídas de conversores. Isso significa que um conjunto de serviços de conversão estará disponível em rede, e será utilizado, por meio de *software*, para realizar a migração. Ferreira, Baptista e Ramalho (2005), sugerem que a “criação de uma rede global de migradores poderá conduzir a uma redução generalizada dos custos de preservação”.

Quanto a atualização de versão, Miguel Ferreira (2009) afirma que essa técnica é possivelmente a mais utilizada pela generalidade dos utilizadores. Na atualização de versão a migração ocorre dentro de uma mesma “família” de forma linear, ou seja, à medida que produtos tecnológicos ganham versões mais atuais, estas vão substituindo as versões anteriores. São os próprios fornecedores desses produtos que se encarregam de fornecer as rotinas necessárias que irão permitir que versões mais recentes leiam versões anteriores, possibilitando inclusive que os arquivos sejam salvos na nova versão (Thibodeau, 2002). Porém, Thibodeau refere que essa migração pode fazer com que o arquivo atual apresente características diferentes do arquivo original, como alterações no formato ou cor. Outro potencial problema é a descontinuidade do produto pelo fornecedor.

Já a normalização, consiste na redução do número de formatos distintos ou mesmo na adoção de um formato padrão para os documentos. Nesse sentido a redução ou adoção de um único formato irá permitir a aplicação de uma dada estratégia de preservação transversalmente a um maior número de objetos (Ferreira, 2009), facilitando, por exemplo, a migração de grandes quantidades de dados. É o formato que “indica a codificação, a sequência, a ordenação, o tamanho e as relações internas que o singularizam e que permitem que um arquivo seja devidamente interpretado” (Ferreira, 2011:42). Portanto, a escolha do formato é fator importante e, de acordo com Ferreira (2009:28), devem ser escolhidos, quando possível, formatos reconhecidos pela comunidade de interesse, baseados em normas internacionais abertas, evitando possíveis complicações relativas aos direitos de autor ou pagamentos de *royalties*. Para além disso, a normalização promove a interoperabilidade entre diferentes sistemas.

1.4.2.5. Pedra de Roseta

Uma abordagem mais recente e ainda em desenvolvimento é a Pedra de Roseta³⁰. Ela opera com base na construção de uma amostra a partir de objetos representativos do formato que se pretende recuperar. Essa amostra irá possibilitar que sejam inferidas regras necessárias para a tradução/conversão do objeto original em formatos atuais (Thibodeau, 2002; Ferreira, 2009). Thibodeau (2002) inclui a Pedra de Roseta como uma das formas de migração e, a National Library of Australia (2003), como uma estratégia de preservação. Já para Ferreira (2009), ela não é propriamente uma estratégia, mas uma ferramenta da Arqueologia Digital³¹

1.5. Preservação digital distribuída

“hoje os principais projetos de bibliotecas e de repositórios digitais estão colocando on-line estoques massivos de recursos informacionais em formato digital. Algumas dessas coleções incluem milhões de objetos e estão sendo planejadas para gerenciarem um número astronômico de informações de toda a natureza” (Sayão e Marcondes, 2009:27).

Preservar esses estoques e proteger o conteúdo informacional para que estejam disponíveis e possam ser acedidas no futuro é uma questão proeminente no contexto das universidades. Skinner e Schultz (2010) consideram que os processos relativos a preservação digital ainda estão nos estágios iniciais, porém, apontam a Preservação Digital Distribuída – PDD³² como modelo de preservação mais promissor até o momento. O foco desse modelo é a distribuição de cópias dos objetos

³⁰ Do inglês *Rosetta Stones*

³¹ “uso de material obsoleto para pesquisa da historia dos inícios da arquitetura computacional, aplicativos periféricos, sistemas operacionais, linguagens de programação, ferramentas e sistemas de gestão eletrônica de documentos” (Márdero Arellano, 2009: 350)

³² Do inglês *Distributed Digital Preservation*

digitais em locais geograficamente distintos, de modo seguro e de forma a garantir o acesso de longo prazo (Souza et al, 2012).

Na PDD a colaboração entre as instituições é fator fundamental. Essa colaboração irá exigir que todas as instituições envolvidas invistam recursos técnicos e organizacionais e que firmem acordos interinstitucionais de longo prazo visando atender às necessidades de preservação para as quais foram criadas (Souza et al, 2012). Os acordos devem prever uma metodologia pré-coordenada de forma que as instituições possam combinar estratégias que garantam a distribuição geográfica para vários locais, visando maximizar medidas de segurança e reduzindo a probabilidade de perda ou comprometimento dos conteúdos preservados Skinner e Schultz (2010).

Skinner e Schultz (2010) sugerem princípios essenciais a serem seguidos em um modelo de preservação digital distribuída:

- ✓ As cópias dos conteúdos devem ser mantidas em locais geograficamente dispersos, em um raio de 120 a 200 Km um do outro;
- ✓ Aos selecionar os locais de preservação, procurar evitar lugares com histórico de catástrofes naturais;
- ✓ Os locais de preservação devem ser distribuídos em diferentes redes de energia;
- ✓ Manter o controle dos locais de preservação sob diferentes administradores de sistemas;
- ✓ As cópias dos conteúdos preservados, em diferentes locais, devem ser avaliadas e comparadas regularmente. Isso irá possibilitar possíveis reparações caso algum conteúdo tenha sido corrompido. E, as ações referentes a esses processos devem ser registradas permanentemente no sistema;
- ✓ O conteúdo deve ser replicado pelo menos duas vezes (três cópias), sendo esse o número mínimo para garantir a segurança das cópias.

Para os autores, independentemente da infraestrutura técnica que possa ser adotada em um modelo de PDD, três processos principais serão realizados:

ingestão/coleta, monitoramento, e recuperação dos conteúdos Skinner e Schultz (2010). Citamos aqui a iniciativa de âmbito internacional que adota esse modelo, a Aliança LOCKSS.

1.5.1 LOCKSS³³

O LOCKSS – *Lots Of Copies Keep Stuff Safe* foi criado em 1998 pela Universidade de Stanford, como produto de um projeto cujo objetivo era auxiliar os bibliotecários na preservação dos periódicos acadêmicos publicados na internet e também a continuidade de acesso a esses documentos. O avanço das ações empreendidas no projeto resultou na formalização de uma iniciativa internacional e colaborativa de preservação digital, estabelecida em 2004 e denominada Aliança LOCKSS.

Essa aliança promove um trabalho colaborativo voltado para o acesso autorizado de cópias digitais dos materiais preservados. Ela fornece uma infraestrutura distribuída e coordenada para o gerenciamento de materiais digitais por meio do LOCKSS, um *software* livre e portanto, de código fonte aberto. A aliança conserva dois tipos de rede: global e privada. Na Rede Global são preservados materiais privados de um conjunto de instituições, enquanto a Rede Privada (Private LOCKSS Network – PLN) se responsabiliza por redes locais, regionais ou temáticas, que preservam o conteúdo de repositórios institucionais, documentação oficial do governo, bases de dados, conteúdo de acesso livre, dentre outros materiais de acesso aberto (Márdero Arellano, 2012).

O LOCKSS atua como um compartilhador P2P (Peer-to-peer), ou seja, de forma descentralizada cada computador conectado a rede assume papel de cliente e servidor ao mesmo tempo, possibilitando assim, o compartilhamento de serviços e dados. Nesse modelo o material é coletado, armazenado e preservado localmente nas instituições participantes de uma PLN, não existindo, portanto, um servidor central. As instituições participantes também são referidas como “nó”, e o termo ‘caixa

³³ Disponível em: <http://www.LOCKSS.org/>

LOCKSS' é empregado para cada *software* LOCKSS instalado em um *hardware* apropriado, em cada nó da rede.

O *software* executa cinco funções principais (LOCKSS, 2015):

- ✓ Ingere conteúdo de *sites* de destino utilizando um rastreador de internet similar aos utilizados pelos motores de busca;
- ✓ Preserva o conteúdo armazenado, comparando-o regularmente com o conteúdo replicado e reparando eventuais incoerências;
- ✓ Fornece páginas web para acesso aos conteúdos preservados, quando o *site* de origem não estiver disponível;
- ✓ Disponibiliza interface web para gerenciamento e monitoramento dos conteúdos e para controlar o acesso ao conteúdo preservado;
- ✓ Migra dinamicamente o conteúdo para novos formatos: quando um navegador web não pode exibir o conteúdo o LOCKSS migra dinamicamente esse conteúdo para um formato mais recente, para exibição.

2. Repositórios digitais de acesso livre

Em repositórios digitais de acesso livre, abordam-se aspectos relativos ao movimento do acesso livre como também questões pertinentes aos repositórios digitais, tais como contexto no qual surgiram, conceitos, tipologia e importância desses repositórios.

2.1. Acesso livre

Na década de 80 surgiu o *Open Source Movement* com o intuito de possibilitar o acesso livre à informação existente acerca dos sistemas de *software* (Castells, 2004). Esse acesso se dava por meio da distribuição do código fonte desses sistemas, o que tornou possível a qualquer pessoa realizar alterações no código, desenvolvimento de novos *software* a partir dele, ou ainda, apenas adequar o *software* de acordo com suas necessidades. (Castells, 2004).

Entre os anos de 1980 e 1990 a comunidade científica viu crescer: o número de publicações científicas nas diversas áreas do conhecimento; difusão e uso de *software* livre; juntamente com a generalização da internet e da web (Castells, 2004). Com isso, criou-se um contexto que propiciou o surgimento das primeiras iniciativas do que mais tarde viria a ser chamado de movimento de acesso livre à informação. Esse movimento buscava a distribuição de forma aberta e a nível mundial da literatura de caráter acadêmico ou científico.

A partir de 1998, diante dessa conjuntura e, em um ambiente de inovações tecnológicas que permitia novas possibilidades, o movimento de acesso livre começa a se firmar com a criação da *Scholarly Publishing and Academic Resources Coalition* - SPARC³⁴ e as ‘Declarações de Independência’³⁵ (Rodrigues, 2004). A consolidação

³⁴ A SPARC “é uma aliança internacional de bibliotecas acadêmicas e de pesquisa que trabalha para criar um sistema aberto de comunicação científica” (<http://www.sparc.arl.org/>)

³⁵ “termo usado pela SPARC, significa a demissão dos editores (científicos) de uma revista, em protesto contra a sua política comercial, e o lançamento de uma revista semelhante junto de um editor mais amigável. Em 1998 a maior parte da comissão editorial do Journal of Academic Librarianship demitiu-se e lançou Portal: Libraries and the Academy toda a

do movimento vem com o surgimento da *Open Archives Initiative* – OAI³⁶ (em 1999), da *Budapest Open Access Initiative* - BOAI³⁷ (em 2001), da *Bethesda Statement on Open Access Publishing*³⁸ (em junho de 2003), e por fim, da *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*³⁹ (em outubro 2003). Essas iniciativas trouxeram importantes contributos, firmando alianças, estabelecendo conceitos e ainda fornecendo instrumentos normativos.

2.2. Repositórios digitais

A OAI visando inicialmente o acesso a arquivos de *e-print* desenvolveu e promoveu padrões para tornar interoperáveis diversos repositórios que operavam de acordo com a proposta dos arquivos abertos, armazenando e tornando possível a recuperação de documentos. O *Open Archives Initiative Protocol for Metadata Harvesting* – OAI-PMH⁴⁰ foi o mecanismo criado para permitir esta interoperabilidade por meio de uma estrutura onde provedores de dados e provedores de serviços interagem para permitir o acesso e a preservação da informação (Figura 3). Nesse modelo, os provedores de dados são os responsáveis pela criação e guarda dos metadados necessários para identificação dos conteúdos armazenados, e os provedores de serviços são responsáveis pela coleta dos metadados (por meio do OAI-PMH), disponibilizando os conteúdos de diferentes repositórios em uma única interface de busca e consulta de informações. Para Rodrigues (2004:28) “a iniciativa OAI e o protocolo OAI-PMH contribuíram para dar maior visibilidade e encorajamento ao movimento do Acesso Livre”.

comissão editorial do *Evolutionary Ecology*, liderada por Michael Rosenzweig, demitiu-se e criou *Evolutionary Ecology Research*” (Rodrigues, 2004).

³⁶ Disponível em: <http://www.openarchives.org/>

³⁷ A *Budapest Open Access Initiative* - BOAI, segundo Rodrigues (2004) “estabeleceu o significado e o âmbito do Acesso Livre e definiu duas estratégias complementares para o promover e atingir”. As estratégias definidas pela BOAI foram a via dourada, com objetivo de disseminar publicações por meio de revistas de acesso livre e sem nenhum custo associado para acesso aos artigos, e a via Verde, objetiva o “auto-arquivo, pelos autores, dos artigos publicados nas revistas ‘tradicionais’ de acesso restrito [...] em repositórios de livre acesso, os Repositórios Institucionais” (Rodrigues, 2004). Disponível em: <http://www.budapestopenaccessinitiative.org/>

³⁸ Disponível em: <http://legacy.earlham.edu/~peters/fos/bethesda.htm>

³⁹ Disponível em: <http://openaccess.mpg.de/286432/Berlin-Declaration>

⁴⁰ Disponível em: <http://www.openarchives.org/pmh/>

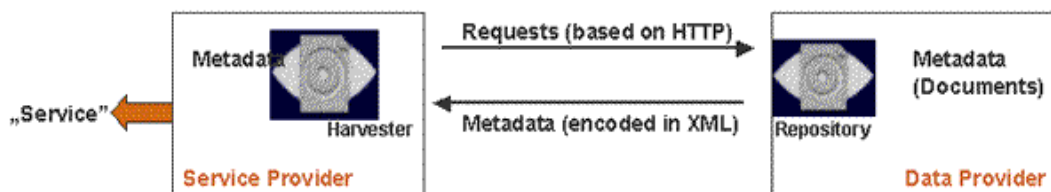


FIGURA 3 - Funcionamento básico do OAI-PMH (Disponível em: <http://www.oaforum.org/tutorial/english/page3.htm>)

Márdero Arellano (2008) aponta que o movimento para criação de repositórios cresceu consideravelmente a partir da disponibilização do pacote de *software* EPrints⁴¹, no final dos anos 2000, pela *University of Southampton* (Inglaterra), juntamente com um contexto propício onde proliferou a banda larga e tecnologias de redes de gigabytes e, ainda, padrões de metadados para descrição de conteúdos em repositórios digitais foram desenvolvidos (Márdero Arellano, 2008). Para além disso, um outro fator importante para adoção dos repositórios digitais foi a chamada ‘crise da comunicação técnico-científica’ caracterizada por fatores como

“alto custo da veiculação de produtos pelas mídias tradicionais; a necessidade de transparência das ações institucionais para subsidiar a avaliação das instituições na luta para a obtenção de recursos; a pressão gerada pela comunidade científica em prol da diminuição da censura, ou proteção excessiva ao acesso às informações; o desenvolvimento acelerado da tecnologia de armazenamento de recursos digitais.” (Dodebei, 2009:92)

A partir de 2002, os repositórios digitais de acesso livre passam então a ser amplamente adotados pelas instituições de ensino e pesquisa como uma alternativa a um contexto de hegemonia dos editores científicos, se tornando uma opção viável e de baixo custo, ampliando as possibilidades de acesso aos conteúdos e se constituindo em uma nova via para a comunicação científica, através da disponibilização de forma livre da literatura científica produzida no âmbito das universidades e instituições de pesquisa (Leite, 2009). Isso permitiu aos utilizadores realizar busca, leitura, *download*, cópia, distribuição ou impressão dessa literatura científica (Leite, 2009). Para Sayão e Marcondes (2009:27), sem a iniciativa OAI e os padrões por ela

⁴¹ Disponível em: <http://www.eprints.org/>

desenvolvidos e promovidos, “teríamos poucas e caras ilhas de acervos digitais que só poderiam ser acedidas via sistemas especiais projetados para cada uma delas”.

De forma simplificada Ferreira (2006:71), define um Repositório Digital – RD como um “sistema de informação responsável por gerir e armazenar material digital”. Márdero Arellano (2008:124) vai além, colocando um RD como um “serviço de armazenamento de objetos digitais que tem a capacidade de manter e gerenciar materiais por longos períodos de tempo e prover seu acesso apropriado”. Um RD compreende, dessa forma, infraestrutura tecnológica e serviço, tendo como finalidade o acesso, uso e o armazenamento de informação.

Rodrigues (2004) distingue dois tipos de repositórios digitais: (1) repositório temático (ou de disciplina), que guarda documentos de uma ou várias áreas de conhecimento; (2) repositório institucional – RI, que lida com a produção intelectual de uma determinada instituição. Para Thomas e McDonald (2007), a classificação dos repositórios não é tão simples, podendo ser problemática uma vez que pode variar de acordo com o contexto no qual está inserido porém, destaca esses dois tipos, Repositórios Temáticos e Repositórios Institucionais, como sendo os principais.

Márdero Arellano (2008:124), aponta um terceiro tipo, Repositório Central, que segundo ele “são provedores de serviços nacionais e internacionais que permitem a reunião de dados coletados tanto de bibliotecas digitais, quanto de repositórios temáticos e repositórios institucionais”. O referido autor ainda aponta uma definição mais alargada para RI, como sendo “sistemas de informação que armazenam, preservam, divulgam e dão acesso à produção intelectual de instituições e comunidades científicas, em formato digital” (Márdero Arellano, 2008:124). Para Lynch (2003), e em acordo com a definição apresentada por Arellano, um RI além de um conjunto de serviço é essencialmente um compromisso institucional com a dispensação dos documentos digitais como também a preservação de longo prazo, a organização, acesso e disponibilização destes documentos.

2.3. Repositórios institucionais

Os Repositórios Institucionais estão diretamente ligados às instituições de ensino superior e foram largamente adotados a partir de 2002, representando uma nova abordagem capaz de influenciar de maneira séria e sistemática as mudanças aceleradas que vêm ocorrendo na produção do saber e na comunicação científica (Lynch, 2003). Segundo Crow (2002), esses repositórios são coleções digitais que capturam e preservam a produção intelectual de uma ou várias comunidades acadêmicas. Para o referido autor um RI pode responder a duas questões estratégicas dessas instituições: (1) melhorar o processo de comunicação ampliando o acesso às pesquisas, reafirmando o controle das universidades sobre a produção acadêmica, aumentando a concorrência e reduzindo o monopólio das publicações tradicionais, reduzindo custos e aumentando a relevância das instituições e bibliotecas; e (2) servir como indicador de qualidade capaz de demonstrar a relevância científica, social e econômica da sua produção científica, aumentando a visibilidade, *status* e valor público.

Dessa forma, como afirma Márdero Arellano (2008:129), um RI nunca deve ser visto como uma iniciativa isolada, ele permite que as bibliotecas possam trabalhar de modo interdisciplinar enquanto os canais formais e informais de comunicação continuam a evoluir. Por meio do uso do RI as universidades contam com um espaço que pode não só receber e abrigar sua produção intelectual, como também suscitar um espaço para o debate crítico, constituindo-se assim num ator político que desempenha fator inédito na comunicação científica. O RI deve portanto, ser entendido como parte de uma “rede ou infraestrutura informacional de um país ou de um domínio institucional destinado a garantir a guarda, preservação a longo prazo e, fundamentalmente, o livre acesso à produção científica de uma dada instituição” (Sayão e Marcondes, 2009:10).

Esses repositórios são, portanto, serviços organicamente vinculados às instituições de forma cumulativa e persistente, comprometidos com a memória, preservação de longo prazo e com o acesso livre a informação científica (Sayão e Marcondes, 2009). Para atender a essa expectativa e para que possa atuar

contribuindo efetivamente com o processo de comunicação, faz-se necessário a adoção de políticas que norteiem o funcionamento do RI desde sua implantação.

No que se refere a inserção dos conteúdos, por natureza os RI são voltados para o auto depósito, colocando a cargo dos membros de cada comunidade a responsabilidade por depositar a produção científica fruto de sua atuação enquanto membro de uma instituição (Márdero Arellano, 2008). Porém, é possível que essa inserção seja realizada de forma mediada, por editores ou por terceiros, devidamente autorizados pelos autores ou ainda por importação de conteúdos já registrados em outros repositórios. Cabe apontar, através das políticas, que caminhos deverão ser seguidos de acordo com a realidade e necessidades específicas de cada comunidade na qual os repositórios se inserem.

2.4. Tecnologias e ferramentas

Como já discutido anteriormente (página 10), as universidades estão direcionando para os repositórios a custódia dos conteúdos produzidos em seu âmbito, com o objetivo de torná-los disponíveis e acessíveis em longo prazo (Lynch, 2003). O uso de ferramentas e mecanismos tecnológicos abertos e interoperáveis em acordo com os princípios do acesso livre e que compreendam aspectos políticos, legais, educacionais e culturais é o que torna isso possível (Sayão e Marcondes, 2009). Atualmente, estão disponíveis diversos pacotes de *software* proprietários e de código fonte aberto, destinados a este fim, capazes de atender satisfatoriamente esta demanda.

Dos diversos *software* disponíveis para criação de RD, são o DSpace⁴², EPrints⁴³ e BePress⁴⁴ os mais utilizados no mundo, de acordo com dados do diretório ROAR – *Registry of Open Access Repositories*⁴⁵. Dos 2632 Repositórios Institucionais ou Departamentais registrados neste diretório, cerca de 45%⁴⁶ faz uso

⁴² Disponível em: <http://www.DSpace.org/>

⁴³ Disponível em: <http://www.eprints.org/>

⁴⁴ Disponível em: <http://digitalcommons.bepress.com/>

⁴⁵ Disponível em: <http://roar.eprints.org/>

⁴⁶ Dados coletados em 07 de janeiro de 2015.

do pacote de *software* DSpace, fazendo deste o mais utilizado para implantação e gestão de RI. No caso do Brasil o diretório registra 150 repositórios digitais, sendo 79 Institucionais ou Departamentais. Destes, 66 utilizam o DSpace.

2.4.1. DSpace

A ferramenta foi desenvolvida, inicialmente, pelo *Massachusetts Institute of Technology* – MIT e laboratórios *Hewlett-Packard* com o intuito de criar repositórios digitais de acesso livre, porém, atualmente é mantido pela DuraSpace⁴⁷ e conta com apoio de uma comunidade mundial⁴⁸.

O DSpace é um *software open-source*, criado com base em padrões internacionalmente aceitos como OAIS e Dublin Core e tem implementado o protocolo OAI-PMH, viabilizando a interoperabilidade. Ele admite o reconhecimento e gerenciamento de variados tipos de materiais, texto, arquivos de som e imagem, por exemplo, e em diferentes formatos de arquivo como PDF, Word, JPEG, MPEG, TIFF, entre outros.

O *software* está estruturado de forma a permitir a organização dos conteúdos por meio de um esquema hierárquico composto por Comunidades, Coleções e Itens organizadas para refletir as estruturas organizacionais das instituições, como também, facilitar a recuperação dos objetos digitais armazenados (Figura 4). Nessa hierarquia, a comunidade é o nível mais alto, podendo ser subdividida em subcomunidades, conseguindo, assim, representar temas ou estruturas organizacionais. Desta forma, as comunidades e subcomunidades refletem apenas a estrutura organizacional do repositório, não contendo diretamente objetos digitais.

⁴⁷ Disponível em: <http://duraspace.org/>

⁴⁸ Disponível em: <https://wiki.duraspace.org/dashboard.action>

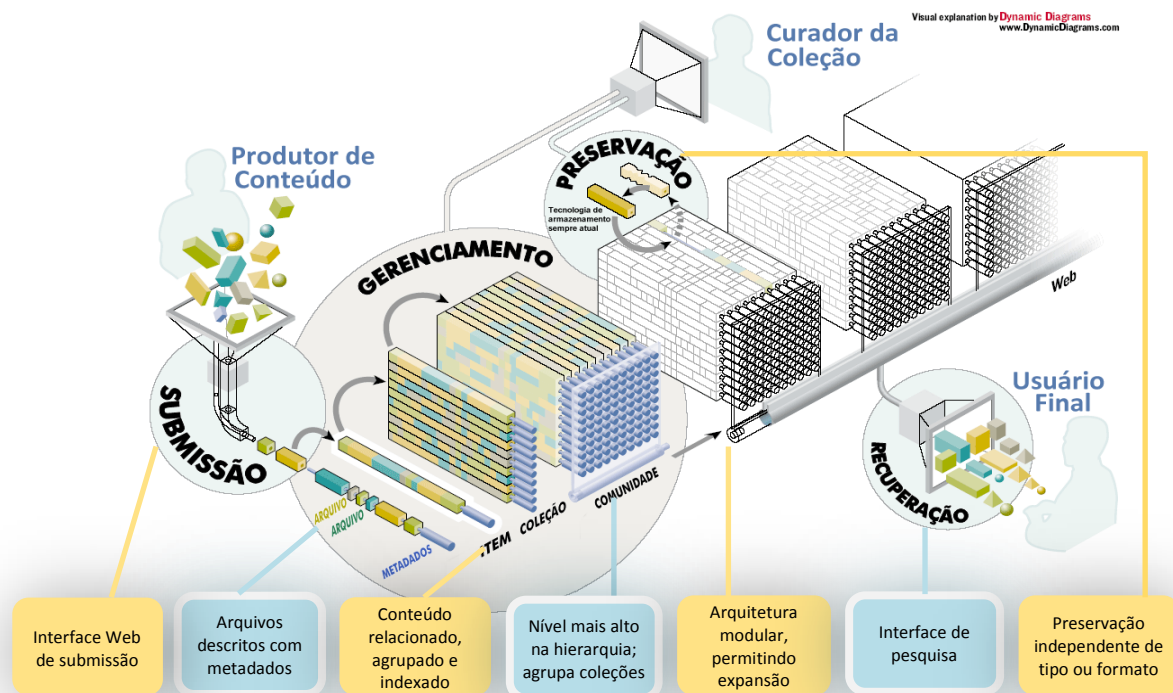


FIGURA 4 – Funcionamento do DSpace (adaptado de <http://www.DSpace.org/introducing>)

A comunidade agrega informação descritiva sobre si mesma e sobre as Coleções. Estas, por sua vez, agregam os Itens ou arquivos, e podem estar vinculadas a uma ou várias coleções. Neste sentido, as coleções refletem a organização dos Itens que compõem o acervo. Um Item é uma unidade informacional formada por dados, que descrevem um documento, e objetos digitais (Shintaku e Meirelles, 2010).

Para cada documento depositado, faz-se necessário a aceitação de uma licença de acesso e uso do conteúdo digital, que é armazenada juntamente com o documento ao qual se refere. Desta forma, cada Item possui mais de um objeto digital. No DSpace os itens são organizados em *bundles*. O *bundle*, ou agrupamento de *bitstream*, é uma estrutura que serve unicamente para agrupar os objetos digitais e o conjunto de metadados.

No que se refere a descrição dos Itens, por padrão o DSpace adota o esquema de metadados Dublin Core, porém, permite que outros esquemas sejam adotados, podendo, ainda, coexistir em um mesmo repositório. O *software* possui três níveis de metadados: descritivo, administrativo e estrutural, que seguem o modelo METS –

*Metadata Encoding & Transmission Standard*⁴⁹. Metadados descritivos são utilizados no nível das comunidades e coleções. Os metadados administrativos agregam informações relativas a preservação, proveniência e dados referentes a política de autorização. E, os estruturais incluem informações relativas a apresentação de um item, ou do *bitstream* dentro de um item, como também as relações entre eles (Diggory e Luyten, 2013).

Para além disso, o DSpace possui mecanismos voltados para a preservação digital de longo prazo, desde a sua concepção. Um desses mecanismos é a identificação persistente dos objetos digitais, que permite a localização, acesso e identificação consistente dos conteúdos por meio de nomes globais exclusivos para cada recurso, que podem ser designados para comunidades, coleções ou itens (Mardero Arellano, 2008). No *software* a identificação persistente é feita com o Handle System®⁵⁰.

Um outro mecanismo utilizado pelo DSpace é o *Checksum Checker* que, objetiva verificar se o conteúdo em um repositório DSpace não foi corrompido ou adulterado. O software atribui a cada item uma identificação única, uma sequência de letras e números, denominada *hash* ou *checksum*. Dessa forma, o mecanismo verifica a integridade dos itens através da verificação do *hash*. O DSpace utiliza ainda o *Storage Resource Broker* – SRB, responsável por armazenar informação sobre os conteúdos, ao nível de *bitstream*. Com esses mecanismos voltados para a preservação digital e longo prazo, o DSpace assegura a conservação de arquivos, possibilitando seu gerenciamento e salvaguarda (DSpace, 2015).

O DSpace, dentre os *software* disponíveis atualmente para a criação e gestão de RD, é o mais utilizado no âmbito das instituições de ensino e pesquisa, conforme discutido anteriormente (página 47) No caso do Brasil, foi amplamente adotado pelas Universidades públicas para a criação de seus respectivos RI, tendo sido por esta razão detalhado em pormenor neste estudo.

⁴⁹ Disponível em: <http://www.loc.gov/standards/mets/mets-home.html>

⁵⁰ Disponível em: <http://www.handle.net/>

2.5. Síntese crítica

O levantamento bibliográfico permitiu identificar os principais conceitos relacionados à área dos repositórios digitais de acesso livre e da preservação digital de longo prazo. Com isso discutiu-se os caminhos, a partir do movimento do acesso livre, que levaram a criação e evolução desses repositórios, seus tipos e finalidades. Baseando-se em dados de diretórios como o ROAR e OpenDOAR, identificou-se o DSpace como sendo um dos principais *software* utilizados para criação dos referidos repositórios, e foram apresentadas as principais características do *software*.

A partir dessa discussão teórica foi possível compreender de forma geral o funcionamento do DSpace e com isso, a adoção, pelo IBICT, do *software* enquanto ferramenta para criação e implantação dos repositórios institucionais nas universidades públicas brasileiras, como é o caso do RI UFRN.

No que se refere à preservação digital, os principais conceitos relativos ao tema, como objeto digital, autenticidade e integridade foram discutidos juntamente com as diversas estratégias para a preservação dos conteúdos, categorizadas em estruturais e operacionais. Nas estratégias estruturais, apresentou-se o modelo de referência OAIS, adotado como padrão para a preservação digital, e os metadados, como necessários para a descrição dos conteúdos e ainda apontados como fundamentais para a garantia de integridade e autenticidades desses conteúdos. Quanto às estratégias operacionais foram citadas as estratégias de preservação da tecnologia, emulação, encapsulamento, migração, refrescamento e Pedra de Roseta. Para além dessas estratégias foi analisado o modelo de preservação digital distribuída e o *software* LOCKSS.

A revisão da literatura possibilitou, portanto, o entendimento das referidas áreas e, também, compreender a perspectiva da preservação digital distribuída como estratégia segura e viável para a preservação digital, e do *software* LOCKSS, que utiliza essa modelo de preservação. Por essa razão, o modelo foi escolhido pelo IBICT para a criação de uma rede brasileira de serviços de preservação digital, a Rede Cariniana, operacionalizada através do *software* LOCKSS.

3. Diagnóstico e análise do contexto institucional de cooperação

No contexto institucional de cooperação caracteriza-se o RI UFRN e a Rede Cariniana, expondo processos e serviços e identificando relações existentes entre ambos. Essa caracterização se deu a partir da coleta de dados realizada por meio de leitura/análise de fontes informacionais relativas ao RI e à Rede Cariniana (Apêndice B), como também com a realização de entrevistas direcionadas aos coordenadores e técnicos (Apêndice A). Apresenta ainda informações relativas aos primeiros testes de integração DSpace - LOCKSS.

3.1. Repositório Institucional da Universidade Federal do Rio Grande do Norte

O RI UFRN foi criado em 2010 com a missão de armazenar, preservar e disponibilizar texto completo de acesso livre da comunidade universitária, especificamente de docentes, técnicos e alunos de pós-graduação (mestrado e doutorado). Como objetivos visa gerir e disseminar essa produção em meio digital, proporcionar visibilidade a essa produção, preservar a memória institucional, e servir como indicador tangível de qualidade e de relevância científica, econômica e social da UFRN.

A criação do repositório se deu a partir do Edital de Chamada FINEP/PCAL/XBDB Nº 002/2009. O edital lançado pelo IBICT teve como objetivo apoiar projetos de implantação de Repositórios Institucionais nas instituições públicas federais, estaduais e municipais de ensino e pesquisa, com o intuito de viabilizar o registro e a disseminação da produção científica dessas instituições, assim como de proporcionar visibilidade a essa produção.

A UFRN enquanto instituição de ensino, pesquisa e extensão participou da seleção promovida através desse edital. Como contemplada, recebeu do IBICT um *kit*

tecnológico⁵¹ composto por um servidor pré-formatado. Para além do *kit*, o IBICT enquanto instituição parceira promoveu treinamentos e disponibilizou conteúdos informacionais para dar suporte ao processo de implantação e gestão dos repositórios.

Em cumprimento aos requisitos e exigências especificadas no edital a UFRN criou o RI e instituiu a Resolução Nº 059/2010-CONSEPE, de 13 de Abril de 2010⁵², que estabeleceu normas sobre a Política Institucional de Informação Técnico-Científica na instituição. Para implantação e manutenção do RI UFRN, a referida resolução determinou a criação de uma Comissão Gestora formada por cinco membros da comunidade, sendo: um representante da Pró-reitora de Pesquisa; um do sistema de bibliotecas da instituição; um da superintendência de comunicação; um da superintendência de informática; e um do departamento de biblioteconomia.

Ressalta-se que a essa comissão atuou na gestão do repositório até meados de 2011⁵³. A partir desse período a Biblioteca Central Zila Mamede – BCZM⁵⁴ como principal unidade de informação da universidade assumiu a gestão e manutenção do repositório. Em 2013 a BCZM criou Setor de Repositórios Digitais – SRD (Anexo A) que passou a responder pela gestão e manutenção do RI, pelo Portal de Periódicos⁵⁵, Biblioteca Digital de Monografia – BDM⁵⁶, e Biblioteca Digital de Teses e Dissertações – BDTD⁵⁷ (anteriormente esses repositórios ficavam sob a responsabilidade do Setor de Automação e Estatística) (Guedes, 2015). Com a desativação da BDTD, em Dezembro de 2014, e dada a fragilidade do *software* e descontinuidade das atualizações, o banco de dados foi migrado para o RI UFRN (Guedes, 2015). Dessa forma, a BDTD passou a ser uma comunidade intitulada Biblioteca Digital de Teses e Dissertações, dentro do RI UFRN (Guedes, 2015).

⁵¹ Descrição do *kit* tecnológico: 01 servidor DELL PowerEdge T300 com processador Xeon Quad Core; 4Gb de memória RAM; 2HDs SATA de 52Gb em RAID 1 (espelhados); sistema operativo baseado em UNIX/LINUX; *software* DSpace 1.5.1 e SEER - Sistema Eletrónico de Editoração de Revistas; pacotes Postgresql, Tomcat 6, Java versão JDK 1.5, Pacotes Java-Ant, Ant-optional e Maven 2, MySQL 5, PHP 5, e Apache 2.1.

⁵² Disponível em: http://repositorio.ufrn.br:8080/jspui/documentos/resolucao_592010_consepe_riufrn.pdf

⁵³ Nesse período, por motivo de mudança na gestão da universidade, alguns membros foram designados para outros cargos e funções. Isso fez com que a gestão do repositório, por recomendação da Reitoria, ficasse sob a responsabilidade direta da BCZM.

⁵⁴ Unidade suplementar da UFRN diretamente subordinada à Reitoria. A BCZM é a unidade central do Sistema de Bibliotecas da UFRN, composto atualmente por 22 bibliotecas. Disponível em: <http://www.sisbi.ufrn.br/bczm/>

⁵⁵ Criado em 2009 o portal abriga as revistas editadas no âmbito da UFRN. Disponível em: <http://www.periodicos.ufrn.br/>

⁵⁶ Criada em maio de 2015 com o objetivo de armazenar e gerenciar os trabalhos de conclusão dos cursos de licenciatura e bacharelado. Disponível em: <http://monografias.ufrn.br:8080/jspui/>

⁵⁷ A BDTD foi criada em 2006 e implantada e gerida por meio do *software* TEDE. Porém, com a fragilidade do *software* e descontinuidade das atualizações, a BDTD foi desativada em Dezembro de 2014 e seu banco de dados foi migrado para o RI. A BDTD armazenava as teses e dissertações defendidas no âmbito da UFRN.

A Resolução estabeleceu ainda, em seu artigo terceiro, que o depósito de informação referentes à produção técnico-científica será registrado no RI pela comunidade universitária, e que seu acesso deverá ser livre nos contextos nacional e internacional. Porém, destaca em seu artigo nono que a BCZM poderá realizar o depósito designado, mediante autorização pelos autores, ou importar dados já disponibilizados em outros repositórios. Hoje cerca de 6% dos conteúdos são depositados no repositório por auto arquivamento, e os demais por depósito designado (Guedes, 2015). Essa produção técnico-científica compreende artigos publicados em periódicos, trabalhos completos apresentados em conferências, dissertações e teses, livro eletrônico, e capítulos de livros.

Como medida de segurança, o servidor que hospeda o RI UFRN foi virtualizado em dezembro de 2014 no datacenter da Superintendência de Informática - SINFO da universidade (Cabral, 2015). Com essa medida foi possível ampliar a capacidade de armazenamento do repositório para 200Gb (podendo ser estendida caso necessário) e garantir a segurança dos dados, uma vez que o datacenter mantém política de *backup*, que determina cópias de segurança diárias, quinzenais, e a replicação do conteúdo do próprio datacenter em outro datacenter (Cabral, 2015). Hoje, o banco de dados do repositório ocupa cerca de 40GB de memória em disco⁵⁸ (Cabral, 2015).

Após a virtualização, ainda em dezembro de 2014, o SRD promoveu a atualização de versão do DSpace, da versão 1.5.1 para a 4.2 com interface web em JSPUI (Java Server Pages User Interface). A opção pela JSPUI partiu da equipe técnica da BCZM por considerar o seu desenvolvimento mais simples e por ter *design* responsivo, o que permite a adequação da interface em diferentes tipos de dispositivos, como computadores e dispositivos móveis, e oferece aos utilizadores uma experiência de navegação similar à do sítio original, independentemente da plataforma (Cabral, 2015). Contudo, o processo de customização do DSpace não está concluído⁵⁹.

A gestão e manutenção do RI UFRN, assim como dos demais repositórios digitais sob a responsabilidade do SRD ficam a cargo de uma equipe de colaboradores

⁵⁸ Dado informado em 28 abril 2015.

⁵⁹ A equipe responsável pela customização está trabalhando na tradução dos emails de resposta do DSpace, customização das planilhas de submissão (pretende-se com isso criar planilhas individuais, de acordo com o tipo de material), e implementação das estatísticas.

formada por de três bibliotecários, um colaborador de apoio técnico e quatro estagiários, e também um analista de sistemas e um técnico de TI, ambos lotados no Setor de Suporte Técnico. O Repositório conta ainda com o apoio da SINFO, responsável por manter o servidor virtual (Cabral, 2015).

3.1.1. Arquitetura de informação

O RI UFRN está organizado de forma a refletir a estrutura organizacional da universidade. Dessa forma, comunidades, subcomunidades e coleções correspondem a (Figura 5):

- ✓ Comunidades – centros, unidades suplementares e unidades acadêmicas especializadas;
- ✓ Subcomunidades – departamentos, cursos e programas de pós-graduação;
- ✓ Coleções – artigos publicados em periódicos, Trabalhos completos apresentados em eventos, Dissertações e Teses defendidas em outras instituições, Livros eletrônicos e Capítulos de livros.

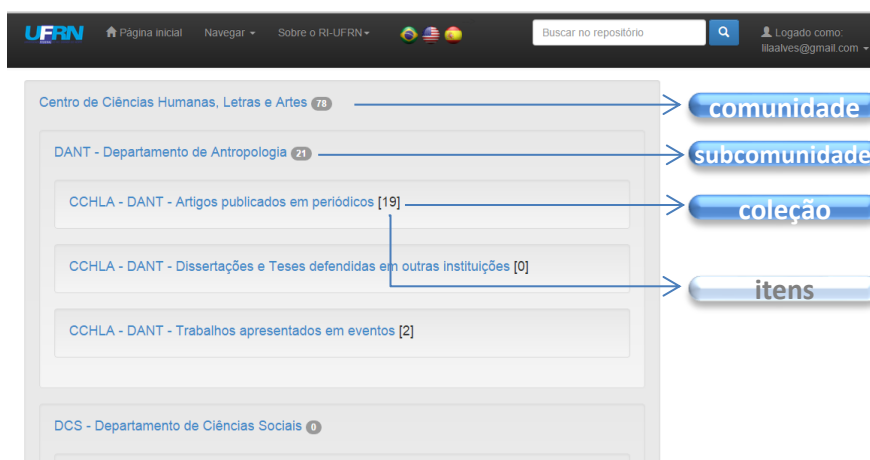


Figura 5 – Organização do RI UFRN

Para que cada coleção seja reconhecida de forma única, siglas foram criadas, identificando a hierarquia de comunidade e subcomunidade, conferindo um padrão de identificação. Assim, a coleção **CCHLA – DANT – Artigos de periódicos** identifica artigos de periódicos cujos autores estão vinculados ao **DANT – Departamento de Antropologia**, que por sua vez é uma subcomunidade do **CCHLA – Centro de Ciências Humanas Letras e Artes** (Figura 5).

A exceção a essa hierarquia é a comunidade Biblioteca Digital de Teses e Dissertações - BDTD (Figura 6), formada a partir da já mencionada migração do banco de dados da antiga BDTD para o repositório. Nesta, a comunidade corresponde a BDTD, as subcomunidades aos programas de pós-graduação, e as coleções separam mestrados e doutorados.



Figura 6 – Comunidade BDTD

3.1.2. Formatos e identificação de arquivos

Para os arquivos relativos aos itens que compõem as coleções não há um padrão quanto aos formatos e à identificação desses arquivos. Por padrão, o DSpace aceita uma ampla variedade de formatos. Contudo, no repositório encontra-se arquivos nos formatos PDF (em diferentes versões) e DOC (.doc e .docx).

Esses arquivos estão identificados de formas distintas. Cada utilizador, no caso do auto arquivamento, atribuiu um nome de acordo com critérios próprios. Esses nomes não sofreram modificação durante o fluxo de aprovação (Figura 8), permanecendo portanto, inalterados. De igual forma, para o depósito designado os arquivos foram nomeados de formas distintas.

No entanto, arquivos inseridos a partir de 2012 na comunidade BDTD encontra-se nomeados de igual forma, sendo o nome composto pelo primeiro nome do autor seguido das iniciais dos sobrenomes e identificação do tipo de documento; exemplos: AlcioCP_DISSERT e LuciaFLGC_TESE.

3.1.3. Depósito e Licenças

Em acordo com a Política Institucional de Informação Técnico-Científica da UFRN, no que se refere ao seu repositório, o depósito da produção técnico-científica dos membros da comunidade acadêmica é realizado de forma designada⁶⁰ e por auto depósito. No depósito designado são os colaboradores do SRD os responsáveis por depositar os conteúdos, que podem ser provenientes dos próprios autores ou de outros repositórios de acesso livre, respeitando os direitos e licenças declarados. Em caso de transferência de direitos de *copyright* a um editor, o conteúdo somente se tornará público a partir da verificação da política de direitos. Para essa verificação é utilizado o serviço considerado de referência nessa área, o SHERPA/RoMEO⁶¹. Com isso, os conteúdos somente se tornam públicos com as devidas licenças, respeitando assim as leis de amparo à produção intelectual.

Quanto as licenças, por padrão o DSpace adota as licenças *Creative Commons*⁶². Contudo, o RI UFRN adota ainda uma Licença de Distribuição Não-Exclusiva (Anexo B) de uso geral, que abrange todos os documentos aceitos pelo repositório. Essa licença concede à UFRN o direito não-exclusivo de reproduzir,

⁶⁰ Atualmente encontra-se em desenvolvimento um plugin de integração SIGAA - Sistema Integrado de Gestão de Atividades Acadêmicas da UFRN - DSpace. Com isso pretende-se recuperar a partir do SIGAA as teses e dissertações de discentes da universidade, que são inseridas obrigatoriamente no SIGAA (Guedes, 2015).

⁶¹ Disponível em: <http://www.sherpa.ac.uk/romeo/>

⁶² Disponível em: <http://creativecommons.org/>

converter, comunicar e/ou distribuir o documento em formato digital ou impresso e em qualquer meio. Com isso, ao submeter um item ao repositório, são expostas duas licenças aos utilizadores, sendo as *Creative Commons* opcionais, e a licença não-exclusiva de adesão obrigatória.

3.1.4. Fluxo de submissão

No DSpace os processos relativos ao fluxo de submissão compreendem três diferentes etapas, depósito (Figura 7), avaliação e revisão. As etapas de avaliação e revisão são opcionais, e formam o fluxo de aprovação. O RI UFRN adota o fluxo completo composto das três etapas, o que implica em maior controle sobre os itens. Dessa forma, para que um item seja disponibilizado necessariamente terá que ser avaliado, e ter seus metadados revisados.

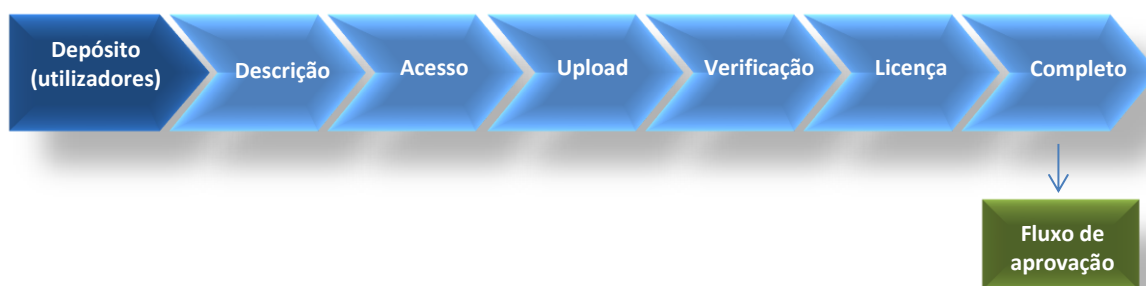


Figura 7 – Depósito de Itens

O fluxo de depósito é considerado simples e o utilizador é conduzido de modo intuitivo através do processo de submissão. Após essa etapa os itens ficam disponíveis para avaliação e revisão. Os bibliotecários do SRD assumem o papel de revisores e avaliadores, sendo, portanto, responsáveis pela revisão final e disponibilização dos itens submetidos ao repositório (Figura 8). Aos bibliotecários, juntamente com o Analista de Sistemas e Técnico de TI, cabe ainda a função de administradores do sistema, sendo assim responsáveis pela gestão e manutenção do repositório como um todo.



Figura 8 – Fluxo de aprovação

Durante a execução do fluxo de aprovação é possível que um item seja rejeitado. Essa opção somente é utilizada quando um item não é adequado a coleção à qual foi submetido, sendo, assim, recomendada a sua nova submissão em coleção adequada, ou se o tipo de documento submetido não estiver de acordo com os tipos aceitos pelo repositório. Os demais itens submetidos passarão por processo de revisão/edição de metadados, e disponibilização para consulta pública.

3.1.5. Metadados

Conforme discutido anteriormente (página 52), o servidor recebido para criação do RI UFRN possuía pré-instalado o DSpace e neste, o esquema de metadados adotado por padrão para os repositórios institucionais brasileiros, o Dublin Core. Já a antiga Biblioteca Digital de Teses e Dissertações – BDTD, fazia uso do Padrão Brasileiro de Metadados para Teses e Dissertações – MTD-BR⁶³, criado e implantado no Brasil pelo IBICT.

Por ocasião da migração do banco de dados da BDTD para o repositório, e visando importar adequadamente as dissertações e teses, fez-se necessário expandir o esquema de metadados utilizado pelo DSpace, inserido neste um conjunto de 18 elementos (Tabela 1). Esse conjunto de foi sugerido e fornecido pelo IBICT, com o intuito de favorecer as instituições, que assim como a UFRN pretendiam migrar as

⁶³ O padrão foi desenvolvido para descrição de teses e dissertações e adotado para as bibliotecas digitais brasileiras de teses e dissertações (BDTD).

teses e dissertações do Sistema de Publicação Eletrônica de Teses e Dissertações – TEDE, para o DSpace.

dc.contributor.authorID	dc.contributor.refereesLattes
dc.contributor.authorLattes	dc.identifier.doi
dc.contributor.advisorID	dc.description.resumo
dc.contributor.advisorLattes	dc.publisher.country
dc.contributor.advisor-co	dc.publisher.department
dc.contributor.advisor-coID	dc.publisher.program
dc.contributor.advisor-coLattes	dc.publisher.initials
dc.contributor.referees	dc.subject.cnpq
dc.contributor.refereesID	dc.description.thumbnail

Tabela 1 – Elementos de extensão para Dublin Core – RI UFRN

O conjunto⁶⁴ foi sugerido como base em elementos do MTD-BR existentes no TEDE, que não possuíam elementos correspondentes a estes no esquema de metadados do DSpace. A inserção desse conjunto, no DSpace, foi necessária para que o processo de migração do banco de dados pudesse ser efetivado sem que houvesse perda quanto aos metadados já existentes na BDTD. Cabe ressaltar que essa adequação não comprometeu o conjunto de elementos básicos do Dublin Core, apenas agregou outros elementos necessários a uma realidade específica, não comprometendo portanto, a interoperabilidade com agregadores nacionais e internacionais.

3.2. Rede Cariniana

O IBICT assume no Brasil a missão de promover a competência, o desenvolvimento de recursos e a infraestrutura de informação em ciência e tecnologia para a produção, socialização e integração do conhecimento científico e tecnológico.

⁶⁴ A agregação desse conjunto não compromete a coleta de dados por agregadores nacionais ou internacionais, uma vez que o conjunto de elementos básicos do Dublin Core, não sofreu nenhum tipo de alteração.

Nesse sentido, e dada a importância estratégica desse conhecimento, o instituto assume a preservação digital de longo prazo desses conteúdos enquanto área estratégica, através da criação e coordenação geral de uma rede nacional de preservação.

A Rede Brasileira de Serviços De Preservação Digital, intitulada Rede Cariniana⁶⁵, objetiva “salvaguardar os registros da ciência, tecnologia e do patrimônio cultural do Brasil”, garantindo o acesso contínuo de longo prazo a esses registros (Mardero Arellano, 2012:87). A Rede Cariniana nasce a partir da necessidade de se criar no IBICT um serviço voltado para a preservação digital de documentos eletrônicos brasileiros. As atividades relacionadas à sua criação tiveram início em 2002 com a formação do Comitê Técnico de Estudos Sobre Preservação de Documentos em Arquivos Digitais. Os estudos realizados pelo comitê deram origem ao projeto geral e à política da rede. Efetivamente, as atividades propriamente ditas de preservação digital foram iniciadas em 2012.

O projeto de implantação baseou-se no modelo de Preservação Digital Distribuída – PDD tendo, portanto, uma infraestrutura descentralizada e recursos de computação distribuída. Dessa forma, as instituições depositárias dos conteúdos eletrônicos participam de uma rede com ambiente tecnológico padronizado que garante a segurança dos dados para que possam, assim, ser preservados e de acesso permanente, e colaboram com a gestão da mesma. À coordenação da Rede Cariniana cabe: disponibilizar *software*, aplicações e ambientes multimídia para que as instituições colecionem, armazenem e promovam o acesso aos conteúdos preservados por meio de cópias autorizadas; e gerir, de forma colaborativa, a Rede como um todo.

A primeira etapa do projeto consistiu em uma análise do LOCKSS visando avaliá-lo como um todo. Após essa etapa, em janeiro de 2013, o IBICT decide utilizar o *software*, e assinou acordo com a Aliança LOCKSS. Com isso, a Rede Cariniana passa a ser operacionalizada através do LOCKSS. Pelo acordo, a Rede irá formar uma *Private LOCKSS Network* – PLN tendo como nó central o *Props Server*, um servidor web localizado na *Stanford University*, responsável pelo fornecimento de dados de

⁶⁵ A Rede tem esse nome em alusão a *Cariniana legalis* ou jequitibá, árvore de grande porte que pode atingir cerca de 50 metros de altura e possuir tronco com até 7 metros de diâmetro. A árvore é símbolo de força e longevidade.

propriedade, distribuição de *plugin*⁶⁶, e informações sobre a PLN. Esse servidor guarda informação sobre cada nó (instituição participante) da rede, agindo assim como um repositório de informações (Souza, 2014a), não interferindo portanto na administração da PLN.

Na segunda etapa do projeto estava prevista a estruturação da PLN de periódicos eletrônicos, disponibilizados na plataforma SEER⁶⁷. Para isso, em 2013 o instituto convidou cinco instituições distribuídas geograficamente em quatro diferentes regiões (norte, nordeste, sudeste e sul), nomeadamente a Universidade de São Paulo – USP, Universidade Federal de Santa Maria – UFSM, Universidade Estadual de Campinas – UNICAMP, Universidade Estadual do Maranhão – UEMA, Universidade Federal da Paraíba – UFPB, para dar início as atividades de preservação de rede. Segundo recomendação da Aliança LOCKSS, cada PLN deve conter no mínimo seis nós. Por esta razão foram convidadas cinco instituições que juntamente com o IBICT formaram a primeira PLN de periódicos (Figura 9).

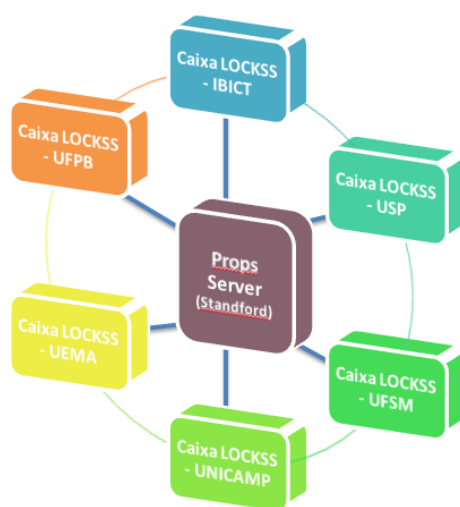


FIGURA 9 – PLN de periódicos eletrônicos
(<http://cariniana.ibict.br/index.php/2014-07-15-17-27-30>)

⁶⁶ “módulos ou extensões de um software. Possibilitam adicionar novas funções ou conectividade com outros software ou hardware” (SOUZA et al, 2012).

⁶⁷ O Open Journal Systems – OJS, desenvolvido pela University of British Columbia, foi taduzido e customizado pelo IBICT e recebeu o nome de Sistema Eletrônico de Editoração de Revistas – SEER.

Cada instituição instalou uma caixa LOCKSS de acordo com recomendações fornecidas pelo IBICT⁶⁸. Nesse modelo, os nós da rede são interligadas, e em cada nó o LOCKSS é executado, coletando dados a serem preservados. A comunicação entre as caixas da PLN se dá via protocolo P2P, tendo como nó central o *Props Server* localizado em *Stanford*.

As caixas existentes na rede são capazes de armazenar, preservar e corrigir danos aos conteúdos, individualmente, e desempenham funções de:

- a) Coleta conteúdo dos websites que ela acessa usando um web crawler similar àqueles usados pelas ferramentas de busca;
- b) Comparação continuada do conteúdo que havia coletado com o mesmo conteúdo coletado por outras caixas e reparação de algumas diferenças;
- c) web proxy ou cache, provendo browsers na comunidade de bibliotecas com acesso ao conteúdo do editor ou ao conteúdo preservado conforme apropriado;
- d) interface administrativa baseada na web, que possibilita ao pessoal da biblioteca apontar novos periódicos para preservação, monitorar o estado dos periódicos que estão sendo preservados e controlar o acesso aos periódicos preservados. (Cariniana, 2013)

Dessa forma, a Rede estruturou inicialmente o serviço de armazenamento dos periódicos eletrônicos das instituições parceiras do projeto que utilizam a plataforma SEER. Para a preservação desses materiais, o LOCKSS se concentra na salvaguarda das cópias dos periódicos, não sendo necessário portanto, um conjunto padronizado de metadados para descrever o material.

Com isso, para a comunicação entre o SEER e o LOCKSS foram necessários apenas três requisitos funcionais: agregar Manifesto LOCKSS⁶⁹, necessário para “declarar conformidade com os processos de coleta e preservação adotados pelo *software*” (IBICT, 2015), e que funciona como uma licença; verificar *links* de referência para acesso a conteúdos suplementares das revistas, certificando-se que foram devidamente enviados; e adicionar o protocolo http ao descrever o código de um link. Para além da observância aos requisitos, cada nó da rede é responsável por indicar os títulos que deverão ser preservados⁷⁰.

⁶⁸ Ver 3.3.2. Componentes tecnológicos

⁶⁹ O SEER, a partir da versão 1.8, passou a ser integrado ao LOCKSS, trazendo assim a opção de adesão ao manifesto já incorporado no sistema, bastando apenas ser habilitado.

⁷⁰ Até Janeiro de 2015, a Rede preservou 603 títulos de periódicos e destes, 5.208 volumes.

O acordo firmado com a Aliança LOCKSS previu fornecimento de suporte técnico do *software* LOCKSS e também a preservação dos livros eletrônicos do Portal do Livro Aberto⁷¹ do IBICT e teses e dissertações depositadas em DSpace, das instituições participantes da rede. Dessa forma, a terceira etapa do projeto (em andamento) visa a ampliação dos serviços da Rede e inclusão desses materiais.

Com esse intuito a equipe do LOCKSS Desenvolveu um *plugin*, com o objetivo de possibilitar a comunicação entre o *software* e o DSpace⁷². A partir desse *plugin* será possível coletar o material armazenado em DSpace para a preservação digital (nessa etapa, apenas os livros do Portal do Livro Aberto e teses e dissertações).

Com essa expansão, e prospectando futuras expansões a Rede Cariniana organizou-se estruturalmente para agregar em sua rede distintas subredes-PLN (Figura 9), de acordo com o tipo de material. Dessas, apenas a PLN de periódicos está efetivada e é formada pelos atuais parceiros integrais (Figura 10).

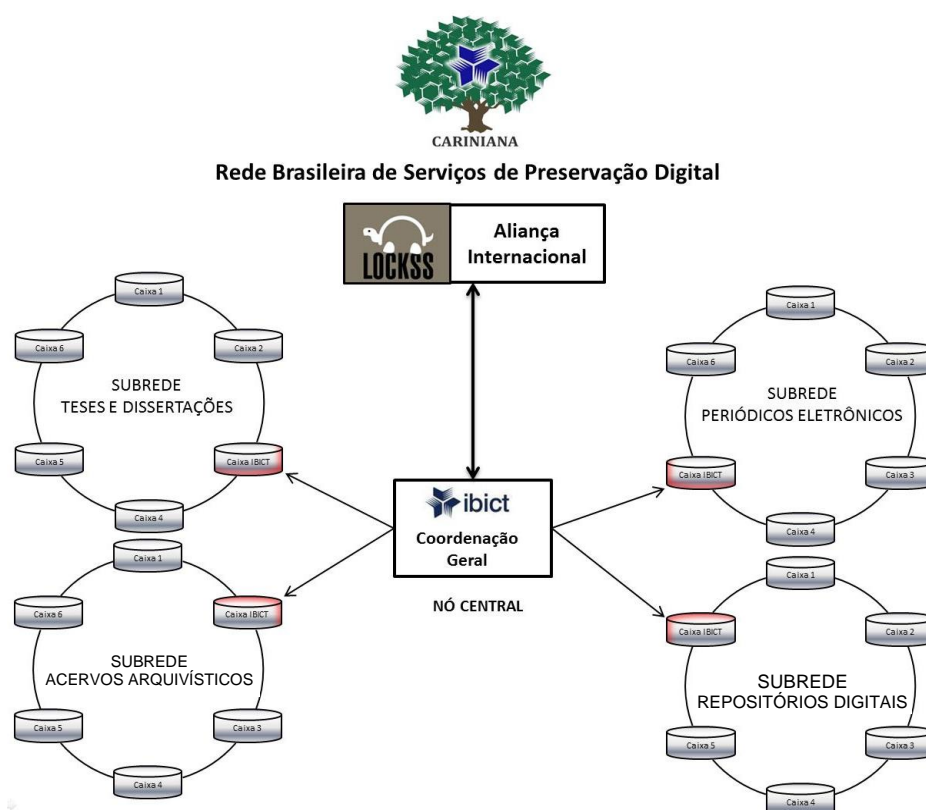


FIGURA 10 – Estrutura Geral Rede Cariniana
(<http://cariniana.ibict.br/index.php/2014-07-15-17-27-30>)

⁷¹ Disponível em: <http://livroaberto.ibict.br/>

⁷² Ver 2.3.3. Testes de comunicação

As demais subredes apresentadas encontram-se em fase de implementação, e assim como na subrede de periódicos, deverão ser formadas a partir dos parceiros integrais, desde que estes possuam matérias correspondentes à subrede e, demonstrem interesse.

Estratégias e mecanismos que agregarão funcionalidades ao LOCKSS, como o plugin de comunicação, estão sendo definidos na terceira e atual etapa de implantação da Rede Cariniana. Com isso, pretende-se facilitar os processos de coleta e armazenamento dos conteúdos armazenados em DSpace, objetos dessa etapa. Por esta razão não há, ainda, documentação regulamentando os processos relativos aos referidos materiais. Assim como ocorreu com os periódicos, esses documentos normativos serão desenvolvidos à medida que a terceira etapa for sendo implementada e seu funcionamento analisado e aperfeiçoado (Márdero Arellano, 2015).

Entretanto, os procedimentos iniciais para a preservação desses materiais são similares aos aplicados a periódicos, no que diz respeito a implantação e gerenciamento das caixas. Portanto, deverão diferir os procedimentos relativos aos processos de coleta dos conteúdos.

Para as próximas etapas, já em fase de estudos, a Rede Cariniana prevê a integração do LOCKSS com outros sistemas eletrônicos de gestão de documentos, nomeadamente, o ICA-AtoM⁷³, Archivematica ⁷⁴, KOHA⁷⁵ e KORA⁷⁶ (Márdero Arellano, 2015).

Paralelamente aos processos propriamente ditos de preservação digital, a rede oferece cursos técnicos e de capacitação, criou a Rede Dríade⁷⁷, vem elaborando juntamente com os parceiros um dicionário eletrônico de preservação digital e um guia de fontes de informação referente a gestão e preservação digital de dados científicos, e ainda, um mapa brasileiro de acervos digitais preserváveis (Cariniana, 2014).

⁷³ Software livre para descrição arquivística. Disponível em: <https://www.ica-atom.org/>

⁷⁴ Sistema de preservação digital. Disponível em: <https://www.archivematica.org/en/>

⁷⁵ Sistema integrado de Bibliotecas, voltado para a gestão de acervos. Disponível em: <http://www.koha.org/>

⁷⁶ Software para criação de repositórios digitais. Disponível em: <http://kora.matrix.msu.edu/>

⁷⁷ Rede de pesquisa sobre estudos e Práticas de Preservação Digital. Disponível em: <http://cariniana.ibict.br/index.php/grupos-de-pesquisas-new>

3.3. Âmbito da cooperação

A cooperação com a Rede Cariniana prevê questões relacionadas com o tipo de parceria que se pretende firmar, a padronização do ambiente tecnológico que irá possibilitar a cooperação e também a efetivação dos processos de ingestão, armazenamento, gestão e acesso, dos conteúdos preservados.

3.3.1 Tipos de parcerias

As primeiras instituições que formaram a PLN foram convidadas para dar início à Rede, permitindo assim que seu funcionamento fosse analisado e aperfeiçoado antes de expandir os serviços da Rede para demais instituições do país. Hoje, a adesão à Rede Cariniana se dá por meio de parcerias com instituições de ensino superior ou que possuam acervos digitais de acesso livre, como também, com usuários e colaboradores individuais envolvidos em projetos acerca da preservação digital. Essas parcerias são efetivadas através de acordo de cooperação, e prevêm os seguintes modelos (IBICT, 2015):

Parceiros Integrais:

- ✓ Requisitos: instituições públicas ou privadas; Compartilhar infraestrutura de armazenamento digital; possuir documentos digitais em formatos adequados para preservação digital.
- ✓ Atribuições: aceitar os termos de funcionamento da Rede, política e acordo; Instalar e monitorar as soluções tecnológicas, de acordo com guia de Interface administrativa do LOCKSS e manual de instalação do LOCKSS; colaborar na administração do serviço; designar representantes para fins de capacitação; participar nos comitês gestores; participar de reuniões técnicas nas

modalidades presenciais e à distância dentro e fora do Brasil; e colaborar na construção dos serviços de preservação digital da Rede Cariniana.

Parceiros Institucionais:

- ✓ Requisitos: participar de algum projeto ligado à preservação digital com vínculo institucional; possuir documentos digitais em formatos adequados para preservação digital.
- ✓ Atribuições: aceitar os termos de funcionamento da Rede Cariniana, política e designação de representantes para fins de capacitação; participar do fórum no site da Rede Cariniana na comunidade que envolve o seu projeto individual; participar de reuniões técnicas nas modalidades presenciais e à distância dentro e fora do Brasil; colaborar na construção dos serviços de preservação digital da Rede Cariniana.

Instituições Usuárias:

- ✓ Requisitos: possuir periódicos na plataforma SEER ou teses, dissertações e livros em repositórios DSpace.
- ✓ Atribuições: aceitar os termos de funcionamento da Rede Cariniana; guia de inclusão de periódicos na rede; atualizar o portal do periódico; apontar conteúdos a serem preservados; monitorar o processo de preservação dos documentos; acompanhar e controlar o acesso aos documentos preservados; informar o tempo de permanência do documento no serviço da Rede Cariniana adotado.

Colaborador individual:

- ✓ Requisitos: participar de algum projeto ou pesquisa ligado à preservação digital.
- ✓ Atribuições: Participar do fórum no site da Rede Cariniana na comunidade que envolve o seu projeto individual; participar de reuniões técnicas nas

modalidades presenciais e à distância dentro e fora do Brasil; e colaborar na construção dos serviços de preservação digital da Rede Cariniana.

Devido ao acordo entre o IBICT e a Aliança LOCKSS, as instituições parceiras são beneficiadas como membros dessa aliança Internacional. Com isso, à equipe da Rede Cariniana cabe o trabalho conjunto com as equipes de cada instituição, colaborando com uso dos serviços oferecidos, como também para desenvolvimento de novos serviços.

O acordo de cooperação com a Rede Cariniana não prevê transferência de recursos financeiros entre as partes, sendo cada instituição responsável pela aquisição de equipamentos, procedimentos técnicos e recursos humanos necessários à integração a Rede Cariniana. Cada acordo tem validade inicial de cinco anos, podendo ser prorrogado mediante Termo Aditivo.

Até abril de 2015, a Rede contabiliza oito parceiros integrais⁷⁸ (Figura 11), um institucional (Jardim Botânico de Brasília – JBB) e diversos parceiros individuais registrados nas comunidades do fórum da Rede Dríade. Essa rede corresponde ao grupo de pesquisa ‘Estudos e Práticas de Preservação Digital’, ligado à Rede Cariniana, possui dez linhas de pesquisa e cerca de cinquenta participantes.

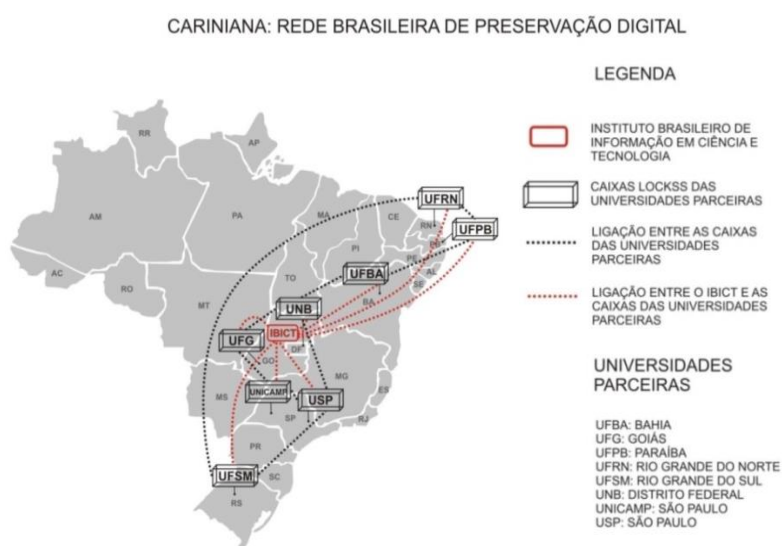


FIGURA 11 – Parceiros integrais da Rede Cariniana

(<http://cariniana.ibict.br/index.php/2014-07-15-17-27-30>)

⁷⁸ Por não possuir equipe técnica necessária, a UEMA não permaneceu na Rede.

Em meados de 2014 a UFRN, objetivando preservar os conteúdos armazenados em seus repositórios digitais, garantir o acesso contínuo de longo prazo a esses conteúdos, e entendendo que para isso, um esforço individual seria dispendioso e moroso, iniciou estudos para aderir à Rede Cariniana. E, em março de 2015, firmou acordo de cooperação técnica, enquanto parceira integral da Rede Cariniana, com o governo brasileiro por intermédio do Ministério da Ciência, Tecnologia e Inovação – MCTI através da unidade de pesquisa IBICT, responsável pela Rede Cariniana.

A partir desse acordo a UFRN seguiu os procedimentos recomendados pela Rede Cariniana, designou equipe responsável, instalou e configurou uma caixa LOCKSS, passando a integrar assim a PLN de periódicos e, conseqüentemente, a salvaguardar as revistas existentes no Portal de Periódicos da UFRN. Até maio de 2015 foram preservadas na Rede 5.467 números de periódicos (ou unidades de armazenamento – Uas). Destas, 52 números, referentes a 11 títulos de períodos, são originários do Portal de Periódicos da UFRN.

3.3.2. Componentes Tecnológicos

A Rede Cariniana, conforme referido anteriormente (2.2.red.cari.página 61), baseia seu funcionamento no LOCKSS. O *software* por sua vez age como um compartilhador P2P, onde cada nó envia e recebe informações em uma hierarquia de rede planificada (Souza, 2014b). Dessa forma, os participantes da PLN compartilham serviços e dados necessitando, portanto, de um ambiente tecnológico padronizado para que os conteúdos possam ser coletados e preservados localmente nos nós de uma PLN.

Para a padronização desse ambiente, a coordenação da Rede Cariniana prevê requisitos mínimos aceitáveis para instalação local das caixas LOCKSS. No que se refere a *hardware*, sugere-se: processador Core 2 Duo; 2GB de memória RAM; e HD de 200GB de memória. Quanto ao *software*, as instituições parceiras deverão

recorrer ao sítio do LOCKSS⁷⁹ para *download* e instalação do sistema operativo CentOS e todos os arquivos necessário para instalação do LOCKSS.

Para o processo de instalação e configuração do software a UFRN assim como as demais instituições definiram alguns parâmetros⁸⁰ que, foram posteriormente informados a equipe do LOCKSS, necessários para tornar as caixas acessíveis ao *Props server*. A partir da instalação e configuração das caixas, e envio desses dados, as instituições passaram a ter efetivo acesso à Rede através da web, via HTTP – *Hypertext Transfer Protocol* ou HTTPS (HTTP seguro) (Cabral, 2015).

As instituições que aderem à Rede Cariniana enquanto parceiras integrais devem ter em consideração o fato da capacidade de armazenamento de disco rígido ser um fator crítico e determinante, uma vez que, irão armazenar não só seus respectivos conteúdos, mas, também, dos demais parceiros integrais. Com base nisso, a UFRN enquanto parceira integral instalou uma caixa LOCKSS em um servidor, com as seguintes configurações (Tabela 2):

Hardware	
Requisitos mínimos Rede Cariniana	Requisitos implementados UFRN
Core 2 Duo	Intel Xeon (R) single core
2 GB de RAM	2 GB de RAM
HD 200GB	HD 1TB

Tabela 2 – Configurações de Servidor

Assim como ocorreu com o RI UFRN, o servidor que hospeda o LOCKSS foi virtualizado e encontra-se na SINFO – Superintendência de Informática, garantindo assim a segurança dos dados e, ainda, possibilitando aumento da capacidade de armazenamento, caso necessário (Cabral, 2015).

⁷⁹ Disponível em: http://www.LOCKSS.org/LOCKSS/Installing_LOCKSS

⁸⁰ Esses parâmetros são relativos a: nome do servidor; endereço de rede; mascara de rede; *gateway*; DNS primário e secundário; *hostname*; endereço de email de administrador LOCKSS; redes que irão acessar o LOCKSS; senha de *root*; usuário e senha para acesso a interface administrativa do LOCKSS, e portas de comunicação (OLIVEIRA, 2013).

Já a caixa LOCKSS do IBICT foi instalada de acordo com as seguintes configurações: CPU de um núcleo com 2 GHz; 2 GB de memória; sistema operativo Linux CentOS 5.1; 500 GB de memória em disco (Souza, 2014b). Do espaço de armazenamento disponibilizado para caixa do IBICT, até 20 de maio de 2015, apenas 40% foi utilizado para a preservação dos conteúdos (Ferrer, 2015).

As caixas instaladas nas atuais instituições parceiras podem coletar diversos tipos de materiais. Contudo, a partir da adesão ou não dessas instituições às demais subredes, cada uma deverá definir a arquitetura de suas respectivas caixas, ou seja, fazer uso de uma única caixa para coleta e salvaguarda de todos os tipos de materiais, ou de caixas distintas, de acordo com o tipo de material. Essa escolha não interfere na administração e funcionamento da rede como um todo. Entretanto, cada instituição deverá considerar questões relativas a capacidade de armazenamento, uma vez que irá salvaguardar seus conteúdos e das demais participantes da Rede Cariniana, como também questões relativas a gestão das caixas⁸¹. Independentemente da arquitetura adotada, o IBICT enquanto coordenador da Rede Cariniana, mantém acesso virtual via protocolo SSH - *Secure Shell*, a todas as caixas que integram a PLN.

3.3.3. Processos de ingestão, armazenamento, gestão e acesso aos conteúdos

A Rede Cariniana, através do LOCKSS, executa quatro funções fundamentais para a preservação digital contemplada pelo Modelo OAIS, a ingestão, armazenamento, gestão e acesso aos conteúdos. Essas funções encontram-se efetivadas para periódicos eletrônicos armazenados em SEER, e deverão funcionar de forma similar para os itens armazenados em DSpace. A exceção deverá ocorrer para o processo de ingestão, uma vez que se trata de materias distintos, necessitando, portanto, de adequações específicas.

As referidas funções são realizadas de forma automática através do *Daemon* LOCKSS, um aplicativo executado em cada caixa LOCKSS. Por meio de uma *interface*

⁸¹ Ver 3.3.3.3. Gestão

Web administradores autorizados podem coletar conteúdo, controlar a forma como o conteúdo será acedido, e monitorar o desempenho do *Daemon*.

3.3.3.1. Ingestão

A ingestão é o processo pelo qual o conteúdo é movido de um *software* (SEER, DSpace dentre outros) para as caixas LOCKSS. Para isso faz-se necessário que o *software* que armazena os conteúdos originais esteja acessível ao LOCKSS. Nesse sentido, *plugins* específicos, para cada *software*, são desenvolvidos pela equipe do LOCKSS de modo a possibilitar a interoperabilidade entre o *software*, como também requisitos mínimos devem ser implementados para essa integração. No caso do SEER, foram necessários três requisitos mínimos, conforme discutido anteriormente (página 54): a implantação do Manifesto LOCKSS; a verificação de *link* de referência; e a adição do protocolo http para descrever *link*, como também um plugin específico.

No caso da ingestão de conteúdos oriundos do SEER (Figura 12), esse processo é precedido pelo registro de informação, junto ao *Props Server*, relativa as Unidades de Arquivamento – Uas (títulos de periódicos juntamente com respectivas URLs – *Uniform Resource Locator*), de cada nó da Rede Cariniana que deverão ser preservados. Essa informação é integrada ao *plugin* de comunicação. A partir disso, as caixas são capazes de rastrear, através de um *crawler*⁸², os dados do Portal que disponibiliza os periódicos, localizando as URLs das UAs para serem coletadas juntamente com metadados correspondentes e com o Manifesto LOCKSS, sendo então preservados da mesma forma que foram disponibilizados na URL original, com suas conexões e demais recursos web (Márdero Arellano, 2012).

⁸² Rastreador semelhante aos utilizados por motores de busca, que segue cadeias de *links* da internet

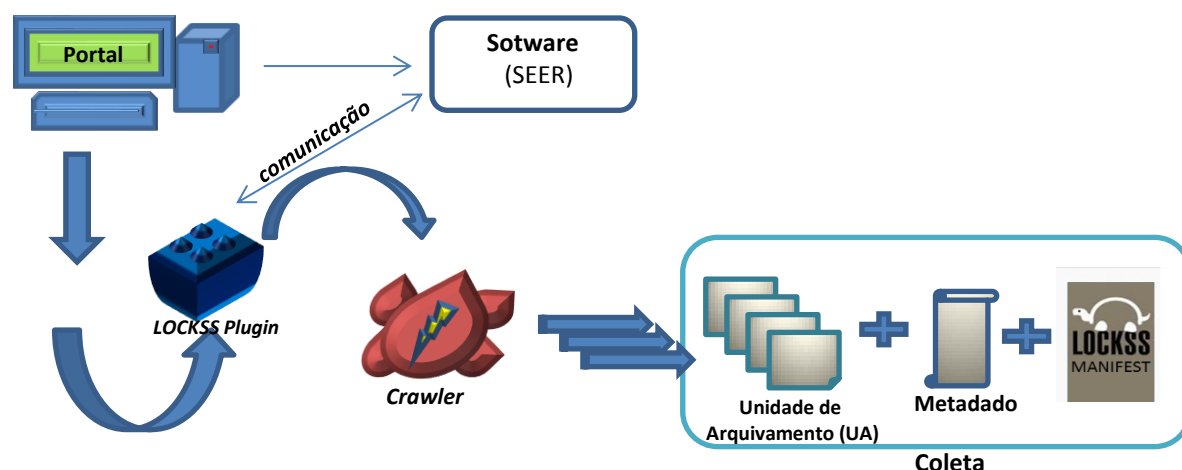


Figura 12 – Processo de ingestão (adaptado de <http://cariniana.ibict.br/index.php/2014-07-15-17-27-30>)

Uma UA corresponde, na PLN de periódicos, a um número de periódico. E cada UA deverá ter associado um manifesto LOCKSS para que possa ser coletado. Esse processo de ingestão de conteúdo não é estaticamente definido, ocorre de forma automática e periódica, sendo único para cada caixa LOCKSS. Assim, a cada novo rastreio, realizado pelo *crawler*, o LOCKSS irá identificar UAs que não constavam em verificação anterior, novos conteúdos, e irá proceder o processo de ingestão dessas unidades.

É possível, através da interface administrativa de cada caixa selecionar UAs para inseri-las manualmente, contudo, esse procedimento é desaconselhado, por oferecer riscos à padronização do conteúdo. Para além disso, trata-se de uma operação local, e portanto, não será replicado na PLN (Souza, 2014b).

Para o DSpace, a partir do desenvolvimento de *plugin* específico⁸³ e da incorporação de requisitos técnicos⁸⁴, o processo procederá de forma similar ao descrito. As UAs serão coletadas juntamente com o conjunto de metadados relativos a cada unidade e licenças atribuídas. Para essa coleta cada nó deverá enviar à coordenação da Rede Cariniana parâmetros que identifiquem que materiais serão preservados. Esses parâmetros irão direcionar o *Daemon* para as coleções específicas, para que os materiais armazenados nelas sejam ingeridos para preservação.

⁸³ Ver 3.3.3. Testes de integração

⁸⁴ Até a conclusão deste estudo, os requisitos técnicos não haviam sido definidos.

3.3.3.2. Armazenamento

Uma vez ingerido, o conteúdo não é apenas armazenado de forma distribuída na PLN, ou seja, replicado nos distintos nós da Rede Cariniana, é ativamente preservado por meio de um processo de verificação de integridade, denominado de *Poll*, que assegura a integridade e autenticidade dos conteúdos. Cada UA ingerida recebe para isso, um hash, que irá identificá-la individualmente e garantir que todas as cópias de uma mesma UA sejam correspondentes.

A etapa inicial do processo, denominada de *votes*, é responsável por eleger/convidar as caixas da PLN para participar do *Poll*. O processo consiste em verificar a integridade dos *bits* das UAs através da comparação do *hash* criado para cada unidade, para garantir que todas as cópias sejam equivalentes. Essa verificação é realizada via P2P entre os nós da Rede. Se durante esse processo de comparação forem identificadas cópias distintas de uma mesma UA, ou seja, com algum tipo de dano, o *daemon* identificará a unidade corrompida. A partir dessa identificação o *Daemon* irá comparar as cópias da UA e determinar a que irá prevalecer, desencadeando a partir daí um processo de re-ingestão de conteúdo, reparando o dano (Skinner e Schultz, 2010; Souza, 2014b).

Status of Poll iOkrZ2VeWJUyjrYoGZteFQaH54=								
Volume: Fragmentum Volume 11								
Type: Proof of Retrievability								
Status: Complete								
Agreement: 100.00%								
Start Time: 13:20:46 03/19/14								
Vote Deadline: 13:51:19 03/19/14								
Duration: 30m34s								
Actual End: 13:51:21 03/19/14								
Total URLs In Vote: 193								
Agreeing URLs: 193								
Quorum: 3								
Peer	Status	Agreement	Agreeing URLs	Disagreeing URLs	Poller-only URLs	Voter-only URLs	PSM State	When
TCP:[150.165.241.6]:9729	Complete	100.00%	53	0	0	0		
TCP:[200.18.45.26]:9729	Complete	100.00%	53	0	0	0		
TCP:[200.137.128.139]:9729	Complete	100.00%	53	0	0	0		
TCP:[200.144.183.66]:9729	Complete	19.17%	37	16	0	140		
TCP:[143.106.108.25]:9729	No Response							
TCP:[164.41.201.17]:9729	No Response							

Figura 13 – Resultado do processo de *Poll* (Souza, 2014b)

A figura 13 apresenta os resultados do processo de *Poll* em uma UA armazenada na Rede Cariniana. O *Poll* foi iniciado a partir da caixa LOCKSS do IBICT. Nesse processo quatro caixas (indicadas com Status Complete) ingeriram conteúdos idênticos. Contudo, uma das caixas apresentou baixo índice de concordância, ou seja, apresentou inconsistência nos dados ingeridos (19,17%). Com isso, o *Daemon* irá reparar o conteúdo realizando nova ingestão, após a comparação dos conteúdos, e posteriormente novo *Poll*. Esse processo ocorre de forma automática e periódica. Dessa forma, a Rede Cariniana garante a integridade e autenticidade dos conteúdos preservados (Souza, 2014b).

3.3.3.3. Gestão

A gestão de cada caixa é possível por meio de uma interface administrativa do LOCKSS (Figura 14), acessível às instituições participantes da Rede Cariniana por processo de autenticação via login e senha. Através dessa interface é possível a cada instituição, monitorar os processos realizados pelo *Daemon*. Ao nó central e à coordenação da Rede Cariniana, cabe o monitoramento de todas as caixas da PLN,

sendo possível saber quem, onde e que processos realizou na interface administrativa de cada caixa.

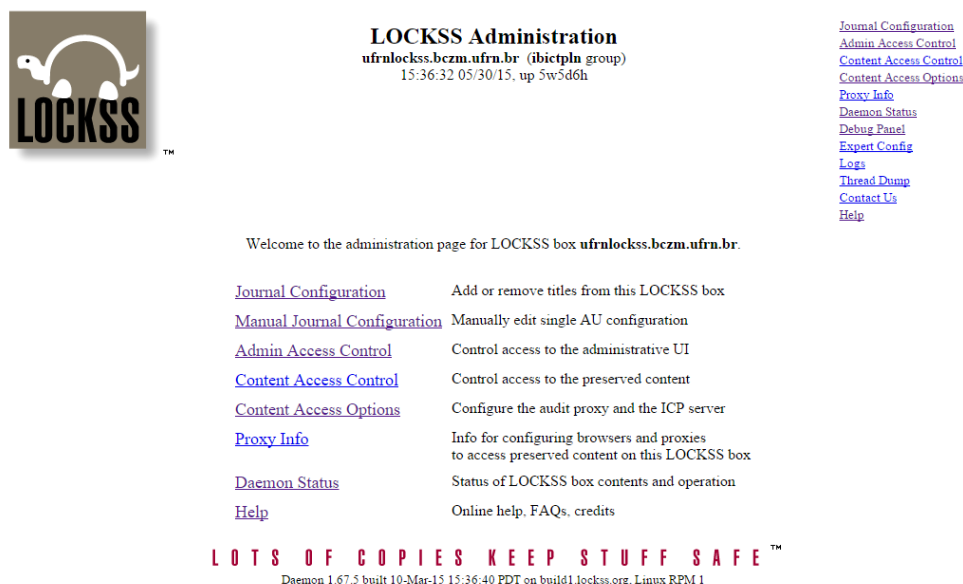


Figura 14 – Interface administrativa da caixa LOCKSS UFRN

A interface apresenta informação referente aos processos de preservação, e oferece algumas funcionalidades que permitem ao administrador de cada caixa, interagir com o sistema, como também oferece comunicação direta com o suporte LOCKSS, através de email. Na interface administrativa é possível (Souza, 2003):

- ✓ Configurar periódicos (Figura 15) – permite visualizar e interagir com os periódicos inseridos para preservação pelo *Props Server*. A funcionalidade permite ainda inserir UAs, remover UA do processo de preservação, desativar temporariamente caso necessário o processo de preservação de UA, reativar um UA, realizar backup de UAs ativas, e restaurar as configurações dos periódicos. Contudo, realizar manualmente algumas dessas ações, pode ferir a padronização do conteúdo. No caso de inserir manualmente uma UA através da interface administrativa, essa ação não será replicada nas outras caixas, o que pode dificultar processos posteriores de preservação (Souza, 2013);



Journal Configuration

ufrnlockss.bczm.ufrn.br (ibictpln group)
07:56:18 07/02/15, up 10w2d22h

Add AUs	Add one or more groups of archival units
Remove AUs	Remove selected archival units
Deactivate AUs	Deactivate selected archival units
Reactivate AUs	Reactivate selected archival units
Backup	Backup cache config to a file on your workstation
Restore	Restore cache config from a file on your workstation
Manual Add/Edit	Add, Edit or Delete an individual AU

LOTS OF COPIES KEEP STUFF SAFE™
Daemon 1.67.5 built 10-Mar-15 15:36:40 PDT on build1.lockss.org, Linux RPM 1

Figura 15 – Configuração de periódico

- ✓ Controle de acesso de administrador (Figura 16) – a interface somente está acessível para IPs (*Internet Protocol*) autorizados. Dessa forma, a funcionalidade permite configurar lista de IP com ou sem permissão de acesso à interface administrativa. Nessa funcionalidade é apresentada uma tabela simples, onde os IPs podem ser registrados;

Admin Access Control
ufrnlockss.bczm.ufrn.br (ibictpln group)
20:04:17 05/30/15, up 5w5d11h

Allow Access ¹	Deny Access ¹
<input type="checkbox"/> 2	
<input type="checkbox"/> /24	
<input type="checkbox"/> 3/26	
<input type="checkbox"/> 124	
<input type="checkbox"/> 8	
<input type="checkbox"/> .16	

Update

Figura 16 – Controle de acesso administrador

- ✓ Controle de acesso de conteúdo – os conteúdos preservados nas caixas somente estarão acessíveis à IPs autorizados. Ainda que um IP tenha

autorização para a visualização da interface administrativa, necessitará ainda de autorização para aceder aos conteúdos. Assim como na funcionalidade anterior, esta possibilita configurar listas de IPs com ou sem permissão de acesso aos conteúdos.

- ✓ Informação de *proxy*⁸⁵ – a funcionalidade (Figura 17) permite recuperar informação relativa a configuração de navegadores e *proxies* para acesso ao conteúdo preservado da caixa LOCKSS na qual a interface está sendo acedida. O LOCKSS pode gerar arquivos PAC⁸⁶ para direcionar esses navegadores para o conteúdo preservado, sendo esta a forma mais simples de configurar um *proxy* para a caixa LOCKSS (SOUZA, 2013).

Proxy Info
ufrnlockss.bczm.ufrn.br (ibictpln group)
22:03:36 05/30/15, up 5w5d13h

This page is used to obtain proxy configuration information for browsers and other proxies, to inform them which URLs should be proxied through this LOCKSS box.

Choose a supported format:

- ☐ **EZproxy config fragment:** Generate text to insert into an EZproxy config file
- ☐ **Generate a dstdomain file for Squid:** Generate a text file that can be used for a Squid "dstdomain" rule
- ☐ **Generate a configuration fragment for Squid:** Generate text to insert into a Squid configuration file
- ☐ **PAC file:** Automatic proxy configuration for browsers. Place the contents of this file on a server for your users to configure their browsers.
- ☐ **Combined PAC file:** PAC file that combines rules in an existing PAC file with the rules for this LOCKSS box.

PAC file URL:

Select preferred source (PAC files only):

- ☒ **Proxy first:** Connect to the LOCKSS proxy first; if no response try the origin server.
- ☐ **Direct first:** Connect to the origin server first; if no response try the LOCKSS proxy.

Figura 17 – Informação de *proxy*

- ✓ Status do *daemon* (Figura 18) – oferece uma visão global do *status* do daemon e apresenta um conjunto de opções de dados sobre os processos executados pelo daemon LOCKSS, em cada caixa. Através desta funcionalidade é possível monitorar os diretamente ligados com a preservação digital, como os processos de crawl, poll e votes, em execução ou já executados. É possível ainda verificar: UA preservadas na caixa; informação sobre dados de comunicação e de configuração da caixa LOCKSS.

⁸⁵ "servidor intermediário que atende a requisições repassando os dados do cliente à frente" (SOUZA, 2013).

⁸⁶ "contém comandos em JavaScript que definem a qual servidor o navegador irá se conectar" (SOUZA, 2013).

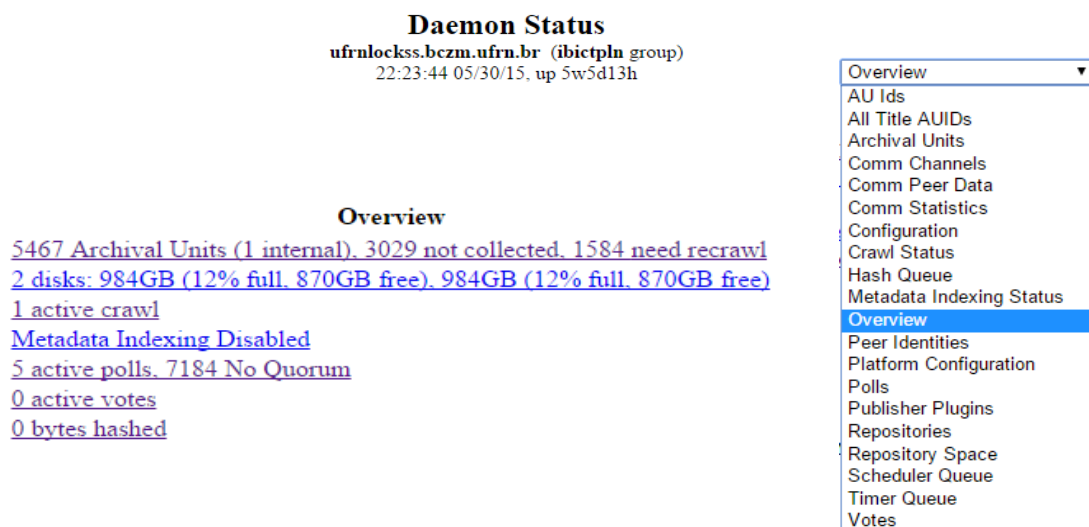


Figura 18 – Status do daemon

- ✓ Painel de *debug* (Figura 19) – através dessa funcionalidade é possível executar comandos manuais no *Daemon* LOCKSS. No entanto, o *daemon* executa as funções de forma automática, assim, não é recomendável utilizar essa funcionalidade, uma vez que irá forçar processos e comprometer as funções.

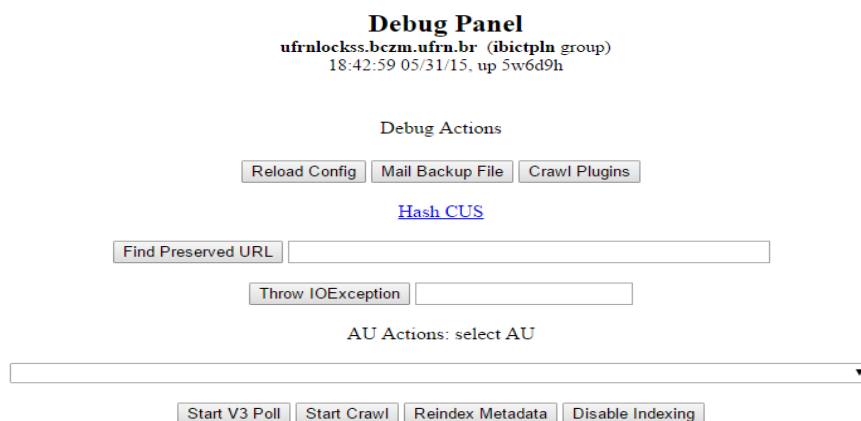


Figura 19 – Painel de Debug

3.3.3.4. Acesso

A partir do modelo distribuído para preservação operacionalizado pelo LOCKSS, em caso de perda parcial ou total dos conteúdos armazenados na PLN por qualquer nó da Rede Cariniana, a restauração desses conteúdos poderá ser realizada a partir de qualquer outra caixa na qual o conteúdo foi replicado. Com isso, a Rede Cariniana assegura aos parceiros a salvaguarda, monitoramento e acesso de longo prazo a esses conteúdos.

Essa recuperação é possível através da interface administrativa, com a funcionalidade **opções de acesso de conteúdo** (Figura 20), que permite configurar como o conteúdo será disponibilizado. Para isso oferece três funcionalidades: opções de conteúdo do servidor; opções do *proxy* cliente; e exportar o conteúdo.

Na funcionalidade **opções de conteúdo do servidor**, é possível especificar definições para gerenciar como a caixa irá fornecer conteúdo. Essa especificação consiste em indicar que portas deverão ser usadas pelo servidor para aceder ao conteúdo, com isso, o servidor de conteúdo irá fornecer uma interface navegável direto ao conteúdo preservado. A funcionalidade **opções do *proxy* cliente**, permite configurar o *crawler* para acessar a rede através de um servidor proxy. E, a funcionalidade **exportar o conteúdo**, que oferece opções para exportação do conteúdo preservado, para o computador que está acessando a interface administrativa. Nesta última, é possível especificar que UA deverá ser exportar e de que forma, sendo possível especificar o tipo de arquivo como ARC, WARC ou ZIP.

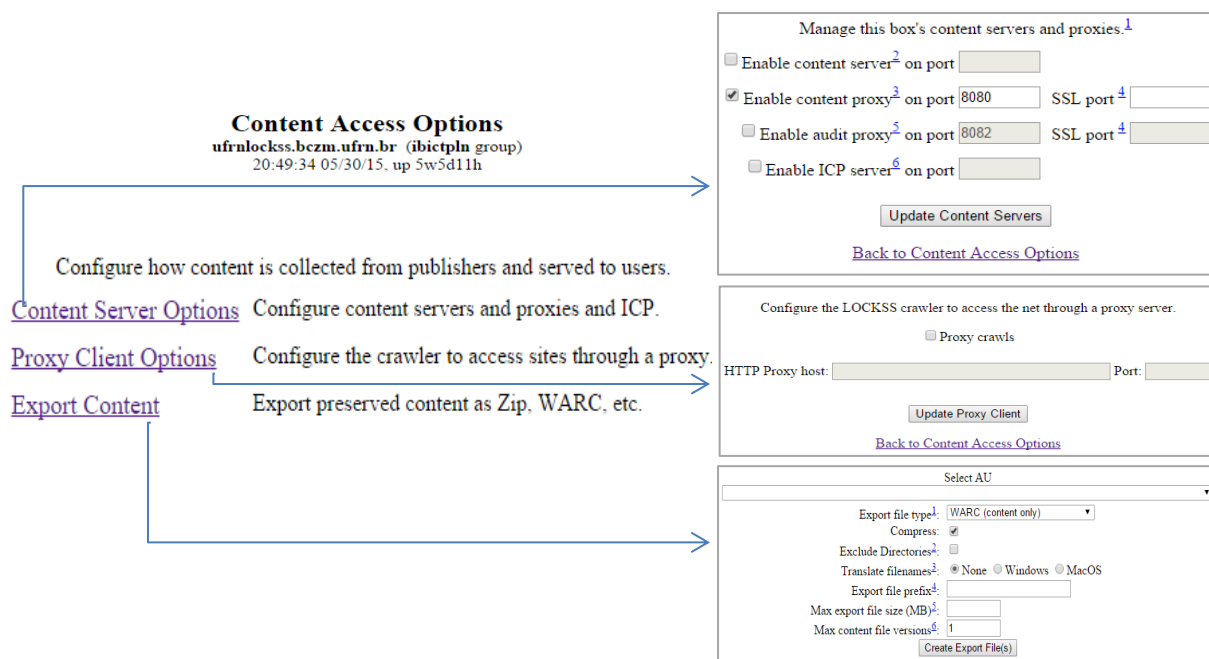


Figura 20 – Opções de acesso de conteúdo

Todavia, apesar dessas funcionalidades estarem disponíveis na interface administrativa de cada caixa, para ter efetivo acesso aos conteúdos preservados é necessário contatar o administrador LOCKSS da Rede Cariniana (IBICT), que irá então abrir exceções nas regras de segurança e enviar parâmetros para que as instituições possam configurar seus portais, no caso da UFRN o Portal de Periódicos Eletrônicos ou RI UFRN (IBICT, 2015).

Em um processo para recuperação de conteúdos, há o risco do navegador não poder processar um formato solicitado, por razões de obsolescência. Contudo, a estratégia utilizada pelo LOCKSS nesse caso será a “migração sobre acesso”. Ou seja, quando um navegador web não conseguir exibir o conteúdo por obsolescência do formato, o LOCKSS irá migrar dinamicamente esse conteúdo para um formato mais recente, criando uma cópia temporária em um formato que o navegador consiga exibir o conteúdo que continuará preservado em seu formato original, e a migração só será executada quando, e se for necessário. (LOCKSS, 2015).

3.3.4. Testes de integração

A integração SEER - LOCKSS já está efetivada. Com isso os processos relativos a preservação dos periódicos armazenados nas instituições parceiras estão sendo realizados satisfatoriamente, e conseqüentemente, os conteúdos estão sendo preservados de forma a garantir o acesso contínuo de longo prazo.

Buscando implementar o serviço de preservação das teses, dissertações e livros eletrônicos do portal do livro aberto armazenados na plataforma DSpace e objetos da terceira etapa da Rede Cariniana, foram realizados os primeiros testes de integração DSpace - LOCKSS entre os dias 23 e 25 de março de 2015 na sede do IBICT em Brasília.

Os testes foram executados pelo , com a participação de técnicos da Rede Cariniana, da UFRN (virtualmente), e da UnB. Para esses primeiros testes foram utilizados os dados do RI UFRN e Repositório da UnB. No caso da UFRN, a equipe técnica clonou o repositório original para a realização dos referidos testes.

A partir de instruções técnicas repassadas pelo especialista da Aliança LOCKSS à equipe técnica da UFRN, um *plugin* de integração foi instalado no DSpace, no repositório clone (Anexo C). Após essa configuração o primeiro teste para coletar as teses e dissertações localizadas na **Comunidade BDTD** foi realizado. Paralelamente ao teste realizado pelo especialista, no repositório clone, o técnico da UnB instalou e testou o *plugin* de integração no repositório UnB.

No entanto, os primeiros testes não obtiveram êxito. Após análise dos resultados os técnicos detectaram uma falha no *plugin* de integração, que gerava uma inconsistência entre a data de disponibilização (correspondente ao campo **dc.date.available**) e a data de criação/publicação (correspondente ao campo **dc.date.issued**), não permitindo assim, que o *daemon* coletasse os conteúdos armazenados nas diversas coleções da comunidade BDTD. Nesse caso, a inconsistência apresentada evidenciou uma necessidade de remodelação dos parâmetros do *plugin*, a fim de orientar precisamente o *daemon* no processo de ingestão para que colete corretamente os conteúdos.

Após os primeiros testes, a equipe da Aliança LOCKSS executou as alterações necessárias no *plugin*, e realizou novos testes de integração DSpace-LOCKSS

utilizando um repositório da *University of Virginia* e obtendo êxito. Assim, novas instruções técnicas para instalação de *plugin* de integração no DSpace, foram enviadas pela equipe LOCKSS à coordenação da Rede Cariniana, para que fossem repassadas às instituições participantes. Em 16 de junho de 2015 essas especificações foram encaminhadas à UFRN para realização de novos testes, não havendo portanto, tempo hábil para realização e publicação de resultados dos novos testes, neste estudo.

No entanto, a partir do teste realizado com o repositório da *University of Virginia* ficou evidenciada a integração entre os dois *software*, de acordo com informações repassadas à coordenação da Rede Cariniana. As etapas seguintes compreenderão a definição das demais instituições que juntamente com a UFRN e o IBICT irão compor a subrede de teses e dissertações, a instalação do *plugin* e início dos processos de ingestão, armazenamento e gestão das teses e dissertações das instituições parceiras, assim como dos livros do Portal do Livro Aberto, do IBICT. A partir disso, seguindo a metodologia adotada pela Rede Cariniana, os documentos normativos referentes a essa subrede serão concluídos e disponibilizados.

4. Conjunto de Requisitos

Este capítulo propõe um conjunto de requisitos voltados para a UFRN e seu repositório institucional, no âmbito do acordo de cooperação com a Rede Cariniana, com o intuito de adequar o repositório e com isso integrar o RI UFRN e a Rede Cariniana.

4.1. Requisitos

Empiricamente, não é possível preservar um conteúdo digital, e sim, a capacidade de reproduzir os resultados. Essa reprodução abrange (1) remontar seus componentes digitais, caso possua mais de um, ou extrair qualquer componente armazenado em um arquivo que contém mais do que um desses componentes; e (2) apresentá-lo na forma adequada (Thibodeau, 2015). Quanto mais informação precisa e não ambígua acerca desses conteúdos e do meio ambiente tecnológico no qual se inserem, maiores as chances de preservação e recuperação dos mesmos, em longo prazo. Dado esse aspecto, a preservação digital precisa começar no início do ciclo de vida desses conteúdos, antes mesmo do desenvolvimento dos processos propriamente ditos, para que seja viável (Skinner e Schultz, 2010).

Um repositório digital por si só é uma ferramenta que oferece um conjunto mínimo de serviços direcionados ao armazenamento, gestão e acesso a conteúdo digital, não sendo, portanto, capaz de preservar esses conteúdos, garantindo acesso contínuo e de longo prazo. Entretanto, a associação desse conjunto a ferramentas, processos e serviços de preservação digital apoiados em normas, modelos e padrões internacionalmente aceitos, com o objetivo de oferecer à comunidade na qual se insere acesso confiável e, de longo prazo a esses conteúdos, torna-o um repositório digital confiável (RLG/OCLC, 2002, p.5).

Para esse fim, estão disponíveis diversas normas, modelos, padrões e programas de certificações reconhecidos e consolidados, como por exemplo: a norma

ISO 16363: 2012⁸⁷ que define práticas recomendadas para avaliar a confiabilidade de repositórios digitais; o modelo de requisitos MoReq⁸⁸, que fornece um conjunto abrangente de requisitos para um sistema de registros; e *The SCAPE Project*⁸⁹, que abordou a preservação a longo prazo de grande escala de coleções de objetos digitais heterogêneos e forneceu diversas ferramentas voltadas para ações de preservação. Para além dessas cita-se, ainda, *The InterPARES Project*, a norma ISO 14721:2012 relativa ao modelo de referência OAIS, as Diretrizes Para Preservação do Patrimônio Digital, fornecidas pela Biblioteca Nacional da Austrália, o PREMIS, já mencionados neste estudo, dentre outras.

Um plano para a certificação para o RI UFRN seria recomendável para estabelecer uma base de confiança. Todavia, o desenvolvimento de processos e a implementação de mecanismos através de práticas confiáveis, comprovadas e disponíveis via normas, modelos e padrões como os citados anteriormente, orientados a esse fim, demanda tempo e custo (OCLC/RLG Working Group on Digital Archive Attributes, 2002).

Embora uma discussão detalhada dessas normas, modelos e padrões esteja além do escopo deste estudo, é possível implementar a curto prazo e com base neles requisitos direcionados ao repositório e à UFRN, no âmbito da cooperação com a Rede Cariniana. Essa cooperação implica em compromissos que devem ser assumidos por ambos os lados (UFRN e IBICT). Dentre outras coisas, no que cabe às instituições parceiras, um desses compromissos diz respeito a adoção de processos e mecanismos que possibilitem o armazenamento de materiais digitais, e à colaboração por parte dos parceiros na construção dos serviços de preservação digital da Rede.

Para os processos já efetivados relativos a PLN de Periódicos, a Rede Cariniana definiu critérios, requisitos mínimos, e elaborou documentação normativa a partir dos testes de integração realizados em conjunto com técnicos da Aliança LOCKSS, diretrizes fornecidas pela aliança, e conhecimento empírico adquirido a partir da efetivação dos processos. Contudo, para a subrede de Teses e Dissertações, estes encontram-se em desenvolvimento. A coordenação da Rede Cariniana iniciou a

⁸⁷ Disponível em: http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

⁸⁸ Disponível em: <http://www.moreq.info/>

⁸⁹ Disponível em: <http://www.scape-project.eu/>

elaboração desses documentos a partir dos testes de integração DSpace-LOCKSS, referidos no capítulo anterior.

Quanto aos procedimentos já existentes para a preservação de periódicos, estima-se que mudanças deverão ocorrer nos processos e procedimentos relativos à ingestão de conteúdos, uma vez que trata de materiais e *software* distintos, permanecendo os demais processos relativos a armazenamento, gestão e acesso, similares. A experiência adquirida com a integração SEER – LOCKSS, mostrou a importância de tratar e padronizar os objetos digitais para que a ingestão possa ser realizada na totalidade, e para que a rede fique operacional. (Souza, 2014b).

Diante disso, buscando apoiar o acordo de cooperação entre a UFRN e o IBICT e entendendo que a preservação digital deve começar bem antes do desenvolvimento dos processos propriamente ditos, sugere-se aqui um conjunto de requisitos mínimos, entendidos neste estudo como necessidade ou parâmetro que deve ser cumprida ou satisfeito (INTERPARES, 2015), a ser implementado a curto prazo e direcionado para a UFRN e seu repositório, no âmbito dessa cooperação.

O conjunto proposto é direcionado para contexto e momento específicos, com o intuito de integrar o RI UFRN e a Rede Cariniana e apoiar o processo de cooperação, e deve ser visto como uma etapa inicial. Com isso, necessidades técnicas mínimas apontadas por iniciativas já consolidadas deverão ser supridas, e por essa razão, o processo poderá ser aplicado a outros repositórios, muito embora não seja esse o objetivo.

4.1.1. Escopo

Na atual etapa a Rede Cariniana prevê a preservação das teses e dissertações armazenadas em DSpace. Para que estes materiais sejam rastreados e consequentemente preservados através do LOCKSS, a UFRN, assim como os demais nós da PLN, terá que enviar à coordenação de rede parâmetros que identifiquem os conteúdos a serem preservados. No caso da PLN de periódicos, foram definidos e

indicados à coordenação da rede os títulos a serem preservados, posteriormente registrados junto ao nó central.

Para os testes de integração descritos no capítulo três, utilizou-se apenas a comunidade BDTD, relativa à produção de discentes dos cursos de mestrado e doutorado da universidade. A escolha se deu por essa produção estar reunida em uma única comunidade, sendo, portanto, de fácil identificação. Inicialmente, o SRD irá optar por preservar apenas essa comunidade.

Porém, o repositório armazena, ainda, teses e dissertações de membros da comunidade universitária, defendidas em outras instituições. Essas teses e dissertações encontram-se distribuídas em diferentes coleções, de acordo com a vinculação do colaborador. Por exemplo, um docente do curso de Biblioteconomia que realizou estudos de doutoramento fora da UFRN, terá sua tese armazenada na coleção CCSA - DECIN⁹⁰ - Dissertações e Teses defendidas em outras instituições. Cabe aqui esclarecer que os conteúdos armazenados no repositório compreendem o período no qual o colaborador está vinculado à universidade.

A decisão acerca do que será preservado deve ser fundamentada e coerente com a missão da instituição e do repositório e deve levar em conta o valor do material (National Library of Australia, 2003). A missão do RI UFRN é armazenar, preservar e disponibilizar texto completo de acesso livre da comunidade universitária, especificamente de docentes, técnicos e alunos de pós-graduação (mestrado e doutorado). Essa produção tem valor a longo prazo e atua enquanto indicador tangível de qualidade e de relevância científica, econômica e social da UFRN, devendo portanto ser preservada.

Nesse sentido, e buscando estar em consonância com a missão do repositório, sugere-se a inclusão das teses e dissertações defendidas em outras instituições, para a preservação na Rede Cariniana. Com isso, os parâmetros para identificação dos conteúdos a serem preservados, terão que ser definidos de forma a direcionar os processos relativos à ingestão de conteúdo no LOCKSS para a identificação e preservação desses materiais.

A determinação do que deverá ser preservado pela Rede cabe a Coordenação do Setor de Repositórios Digitais – SRD. E, os parâmetros que irão direcionar o

⁹⁰ CCSA – Centro de Ciências Sociais Aplicadas; DCIN – Departamento de Ciência da Informação.

Daemon LOCKSS para a coleta dos conteúdos deverão ser especificados pela equipe de TI da BCZM.

4.1.2. Metadados

A preservação digital depende em grande parte da atribuição eficaz de metadados descritivos, administrativos e estruturais, que descrevem, explicam, localizam, ou de outra forma tornam mais fácil a recuperação, uso ou gerenciamento de recursos de informação. Esse requisito trata unicamente de metadados descritivos. Para esse tipo de metadado por padrão o Dspace utiliza o Dublin Core, contudo permite a adoção de outros padrões. Em um modelo de PDD a padronização é fator importante para que os processos relativos a ingestão, armazenamento, gerência e acesso, possam ser realizados satisfatoriamente. Por essa razão, recomenda-se a padronização do conjunto de metadados *Dublin Core*, utilizado para as teses e dissertação. Essa padronização deve ocorrer a nível de elementos atribuídos (formulário integral) e de elementos obrigatórios a serem preenchidos para publicação no repositório (Tabela 4).

4.1.2.1. Padronização dos Formulários

Por ocasião da migração do banco de dados da antiga BDTD para o RI UFRN, e visando importar adequadamente os conteúdos e todos os metadados já existentes no TEDE, a tabela original do DSpace foi estendida. Foram acrescentados à ela, 18 elementos referidos na Tabela 1 (página 60). Com isso, o repositório passou a ter dois conjuntos distintos de metadados para o mesmo tipo de material (teses e dissertações). Em outras palavras, a **Comunidade BDTD** possui um conjunto de metadados distinto do conjunto especificado para as coleções **Dissertação e Teses defendidas em outras instituições**.

Através do DSpace é possível customizar os formulários para descrição dos metadados de acordo com o tipo de material, para que cada um apresente os elementos necessários, evitando assim que campos que não serão utilizados por determinado material fiquem visíveis na tela externa, para os utilizadores.

Nessa perspectiva, e partindo do pressuposto de que todas as teses e dissertações serão preservadas, conforme sugerido no requisito Escopo, sugere-se a padronização do conjunto de metadados para as teses e dissertações (Tabela 3). Os elementos propostos foram determinados com base na tabela de elementos (opcionais e obrigatórios) utilizada no TEDE e em elementos *Dublin Core* utilizados para a descrição nas coleções **Dissertação e Teses defendidas em outras instituições**.

Para essa adequação a equipe de TI terá que configurar os formulários internos do DSpace, agregando os 18 elementos (especificados na Tabela 1:52) incorporados na comunidade BDTD. No entanto, entendendo que uma alimentação retroativa desses elementos pode não ser viável, por demandar significativo tempo e recursos, sugere-se essa adequação, somente, para os novos registros.

	Metadado	Etiqueta (visualização do utilizador)
1	<i>dc.contributor.author</i>	Autor
2	<i>dc.contributor.authorID</i>	CPF do autor
3	<i>dc.contributor.authorLattes</i>	Endereço Lattes do autor
4	<i>dc.contributor.advisor</i>	Orientador
5	<i>dc.contributor.advisorID</i>	CPF do Orientador
6	<i>dc.contributor.advisorLattes</i>	Endereço do Lattes do orientador
7	<i>dc.contributor.advisor-co</i>	Co-orientador
8	<i>dc.contributor.advisor-coID</i>	CPF do co-orientador
9	<i>dc.contributor.advisor-coLattes</i>	Endereço do lattes do Co-orientador
10	<i>dc.contributor.referees</i>	Membro da Banca
11	<i>dc.contributor.refereesID</i>	CPF do membro da Banca
12	<i>dc.contributor.refereesLattes</i>	Endereço do Lattes do membro da banca
13	<i>dc.description.resumo</i>	Resumo
14	<i>dc.description.abstract</i>	Resumo em língua estrangeira
15	<i>dc.description.sponsorship</i>	Agência financiadora
16	<i>dc.description.embargo</i>	Data para disponibilização pública

17	<i>dc.date.issued</i>	Data de defesa
18	<i>dc.publisher</i>	Editor
19	<i>dc.publisher.country</i>	País
20	<i>dc.publisher.department</i>	Departamento
21	<i>dc.publisher.program</i>	Programa de Pós-Graduação
22	<i>dc.publisher.initials</i>	Universidade
23	<i>dc.type</i>	Tipo do documento
24	<i>dc.title</i>	Título
25	<i>dc.title.alternative</i>	Título em língua estrangeira
26	<i>dc.language.iso</i>	Idioma
27	<i>dc.subject</i>	Palavras-chave
28	<i>dc.subject.cnpq</i>	Área do conhecimento CNPQ
29	<i>dc.relation</i>	Relação com outros documentos
30	<i>dc.relation.references</i>	Referências citadas pelo autor
31	<i>dc.description.thumbnail</i>	Endereço eletrônico (handle)
32	<i>dc.description.provenance</i>	Proveniência
33	<i>dc.date.available</i>	Data de disponibilização
34	<i>dc.date.accesioned</i>	Data de depósito
35	<i>dc.identifier.citation</i>	Referência do documento
36	<i>dc.rights</i>	Direitos de acesso
37	<i>dc.format</i>	Formato do arquivo

Tabela 3 – Elementos para descrição das Teses e Dissertações

4.1.2.2. Elementos obrigatórios

Atualmente, o formulário disponibilizado aos utilizadores para o depósito de conteúdos no DSpace, aponta apenas dois elementos como obrigatórios: *dc.title*; e *dc.date.issued*. Por essa razão, além do conjunto padronizado de metadados, sugere-se ainda um conjunto de metadados mínimos, obrigatórios para o depósito dos conteúdos. A tabela 4 apresenta o conjunto padronizado indicando quais metadados seriam de preenchimento obrigatório (marcados em preto) e opcional (marcados em azul), sendo os seis últimos requisitos (marcados em vermelho) de preenchimento automático pelo DSpace, e se são campos que permitem ou não repetição de dados.

Na figura, a coluna Etiqueta, corresponde a nomeação dos elementos para visualização externa, para os utilizadores.

	Metadado <i>Dublin Core</i>	Etiqueta (visualização do utilizador)	Obrigatoriedade	Repetição
1	<i>dc.contributor.author</i>	Autor	SIM	SIM
2	<i>dc.contributor.authorID</i>	CPF do autor	SIM	NÃO
3	<i>dc.contributor.authorLattes</i>	Endereço Lattes do autor	NÃO	NÃO
4	<i>dc.contributor.advisor</i>	Orientador	SIM	NÃO
5	<i>dc.contributor.advisorID</i>	CPF do Orientador	SIM	NÃO
6	<i>dc.contributor.advisorLattes</i>	Endereço do Lattes do orientador	NÃO	NÃO
7	<i>dc.contributor.advisor-co</i>	Co-orientador	NÃO	NÃO
8	<i>dc.contributor.advisor-coID</i>	CPF do co-orientador	NÃO	NÃO
9	<i>dc.contributor.advisor-coLattes</i>	Endereço do lattes do Co-orientador	NÃO	NÃO
10	<i>dc.contributor.referees</i>	Membro da Banca	SIM	SIM
11	<i>dc.contributor.refereesID</i>	CPF do membro da Banca	NÃO	NÃO
12	<i>dc.contributor.refereesLattes</i>	Endereço do Lattes do membro da banca	NÃO	NÃO
13	<i>dc.description.resumo</i>	Resumo	SIM	NÃO
14	<i>dc.description.abstract</i>	Resumo em língua estrangeira	NÃO	SIM
15	<i>dc.description.sponsorship</i>	Agência financiadora	NÃO	SIM
16	<i>dc.description.embargo</i>	Data para disponibilização pública	NÃO	NÃO
17	<i>dc.date.issued</i>	Data de defesa	SIM	NÃO
18	<i>dc.publisher</i>	Editor	SIM	NÃO
19	<i>dc.publisher.country</i>	País	NÃO	NÃO
20	<i>dc.publisher.department</i>	Departamento	SIM	NÃO
21	<i>dc.publisher.program</i>	Programa de Pós-Graduação	SIM	NÃO
22	<i>dc.publisher.initials</i>	Universidade	SIM	NÃO
23	<i>dc.type</i>	Tipo do documento	SIM	NÃO
24	<i>dc.title</i>	Título	SIM	NÃO
25	<i>dc.title.alternative</i>	Título em língua estrangeira	NÃO	SIM
26	<i>dc.language.iso</i>	Idioma	NÃO	NÃO
27	<i>dc.subject</i>	Palavras-chave	SIM	SIM
28	<i>dc.subject.cnpq</i>	Área do conhecimento CNPQ	NÃO	NÃO
29	<i>dc.relation</i>	Relação com outros documentos	NÃO	SIM
30	<i>dc.relation.references</i>	Referências citadas pelo autor	NÃO	NÃO
31	<i>dc.description.thumbnail</i>	Endereço eletrônico (handle)	SIM	NÃO

32	<i>dc.description.provenance</i>	Proveniência	SIM	NÃO
33	<i>dc.date.avaiable</i>	Data de disponibilização	SIM	NÃO
34	<i>dc.date.acessioned</i>	Data de depósito	SIM	NÃO
35	<i>dc.identifier.citation</i>	Referência do documento	SIM	NÃO
36	<i>dc.rights</i>	Direitos de acesso	SIM	NÃO
37	<i>dc.format</i>	Formato do arquivo	SIM	SIM

Tabela 4 – Elementos obrigatórios para descrição das Teses e Dissertações

Caberá a Coordenação do SRD analisar a viabilidade do conjunto proposto, e a equipe de TI da BCZM a customização dos formulários internos e externos do DSpace.

4.1.3. Formato

O formato de um arquivo é uma característica chave de um objeto digital, que orienta a sua utilização. Ele codifica informações em padrões que só podem ser processados e interpretados por combinações específicas de *hardware* e *software*. Atualmente há uma vasta e crescente diversidade de formatos disponíveis no domínio digital, mas, nenhum desses atuais formatos se encontra em risco de se tornar obsoleto. Contudo, é possível não haver maneira de interpretá-los no futuro (Ruusalepp et al, 2012; Stanford University Libraries, 2015).

Para que um conteúdo esteja acessível em longo prazo, é necessário manter informação técnica detalhada sobre os formatos de arquivo utilizados. Com isso, a minimização do número de formatos envolvidos em um sistema simplifica e torna mais prático o gerenciamento de grandes coleções de registros eletrônicos. (Brown, 2008). Há um entendimento quanto as características consideradas necessárias para que um formato seja utilizado com fins de preservação digital de longo prazo (National Library of Australia, 2003; Brown, 2008; Ruusalepp et al, 2012; Stanford University Libraries, 2015): devem ser normalizados; abertos, com especificações

técnicas disponíveis no domínio público, ou que possam ser lidos através de *software* de código fonte aberto; e comumente utilizado em um determinado domínio, o que pode assegurar atualização contínua, e ainda evitar que o sistema seja colocado em uma posição isolada, e talvez não seja interoperável com organizações similares.

Em uma avaliação desenvolvida pela Biblioteca Nacional da Holanda no âmbito do projeto e-Depot⁹¹, foi avaliada a adequação de formatos de arquivo para a preservação de longo prazo com base em critérios como abertura, adoção, complexidade, mecanismos de proteção técnica, auto-documentação, robustez e dependências. Como resultado a avaliação sugeriu a normalização de documentos no formato MS Office e WordPerfect, para um formato de arquivo padrão, o PDF / A. (Rog e Van Wijk, 2015)

O PDF/A⁹² foi criado, pela Adobe⁹³, a partir de uma necessidade de desenvolver um padrão estável, internacionalmente aceito para arquivamento de longo prazo, e construído sobre as especificações PDF⁹⁴. Por natureza, o padrão em si é extremamente técnico e requer, para ser compreendido, especialistas com conhecimento em linguagens de descrição de página (Oettler, 2013; PDF/A – Archiving, 2015). Entretanto, suas principais características são:

- ✓ Formato normalizado, padrão ISO 19005 - Gestão de documentos - formato de arquivo de documento eletrônico para preservação a longo prazo;
- ✓ Incorpora metadados padronizados diretamente no documento;
- ✓ É pesquisável, permitindo processos de reconhecimento de texto;
- ✓ É auto-suficiente: todos os elementos (fontes, perfis de cor, etc.) necessários para uma reprodução autêntica estão incluídos no arquivo;
- ✓ Possui algoritmos de compressão, poupando espaço de arquivamento;
- ✓ Possui mecanismo de proteção técnica, sendo compatível com sistemas de assinaturas digitais.

⁹¹ Projeto direcionado para processamento e armazenamento de longo prazo de objetos digitais. Disponível em:

<https://www.kb.nl/en/organisation/research-expertise/long-term-usability-of-digital-resources/the-e-depot-project-cycle>.

⁹² Possui atualmente, três versões: PDF / A-1 (garante reprodução confiável da aparência visual do documento); PDF / A-2 (permite característica como Compressão JPEG2000; Transparência; Conteúdo Opcional, ou camadas; Assinatura Digital Avançada); e PDF / A-3 (permitir a incorporação de anexos). PDF/A – Archiving (2015).

⁹³ Disponível em: <http://www.adobe.com/>

⁹⁴ Diretamente a partir do OFFICE já é possível a criação de documentos PDF / A.

No RI UFRN os documentos estão arquivados nos formatos PDF (em diferentes versões) e DOC (.doc e .docx). O DSpace, assim como o LOCKSS, permitem que diferentes formatos sejam armazenados. Entretanto, diante da importância do formato para a preservação dos arquivos e com o intuito de normalizar o formato utilizado no RI UFRN para teses e dissertações, sugere-se a adoção do PDF/A que, irá possibilitar a aplicação de estratégias e ações de forma transversal a um maior número de objetos.

A autorização quanto a adoção e normalização do formato cabe à Coordenação do SRD. A implementação do formato caberá à equipe de TI, uma vez que é necessário conhecimento técnico específico para compreender plenamente os requisitos do PDF/A. A coordenação do SRD poderá optar por implantar o formato para os novos arquivos a serem armazenados, ou ainda pela normalização (adoção de um formato padrão) para as teses e dissertações já existente. A normalização é possível uma vez que os formatos existentes no repositório podem ser convertidos para PDF/A.

4.1.4. Nomeação de arquivos

Assim como nas URI, os arquivos devem ser nomeados de forma coerente e não ambígua, aumentando as possibilidades de identificação dos objetos (National Library of Australia, 2003). Atualmente, independentemente da comunidade ou coleção os arquivos no DSpace encontram-se nomeados de diferentes formas, havendo um padrão apenas para a Comunidade BDTD e para os arquivos armazenados a partir de 2012 nessa comunidade. Esse padrão utiliza o primeiro nome do autor seguido das iniciais dos sobrenomes mais o tipo de documento, por exemplo: AlbaO_TESE.

Considerando que o repositório não possui um prefixo *handle* válido, e que o nome dos arquivos é registrado como dado relativo à proveniência do conteúdo, em **dc.description.provenance**, sugere-se como forma de identificar de modo único e não ambíguo os arquivos, nomeá-los com base em palavras do título, data de

publicação e tipo de material, como por exemplo: requisitos-preserv-digital-2015_DISSERT. Para essa nomeação, caracteres especiais devem ser evitados. A implantação desse requisito cabe à Coordenação do SRD, que poderá optar por alterar a nomeação dos arquivos já existentes (a médio e longo prazo), ou adotar esta regra somente a partir dos novos registros.

4.1.5. Arquitetura caixa LOCKSS

Cada instituição parceira da Rede Cariniana instalou uma caixa LOCKSS, que atualmente preserva apenas periódicos eletrônicos. Ocorre que, através da interface administrativa, não é possível visualizar individualmente o conteúdo de cada caixa da UFRN. O que se vê são listas e tabelas com todo o conteúdo reunido, de todas as instituições participantes. Esse conteúdo é exibido em ordem alfabética ou aleatória, dependendo da funcionalidade visualizada.

Para o caso do *daemon* não conseguir efetivar a coleta de alguma UA por razões técnicas diversas, como por exemplo a não habilitação do manifesto LOCKSS no caso do SEER, será possível, através da interface administrativa (funcionalidade: Status do *daemon*), identificar situações como a indicada na Figura 21. Nela se pode verificar que o volume 99 do periódico **Fórum Linguístico** não pôde ser coletado (Status: *No permission from publisher*) por não ter sido habilitado o manifesto LOCKSS, impedindo que o *daemon* realizasse a ingestão.



Daemon Status

ufnlockss.bczm.ufrr.br (ibictpn group)
10:09:10 06/09/15, up 7w1d1h

[Journal Configuration](#)
[Admin Access Control](#)
[Content Access Control](#)
[Content Access Options](#)
[Proxy Info](#)
[Daemon Status](#)
[Debug Panel](#)
[Expert Config](#)
[Logs](#)
[Thread Dump](#)
[Contact Us](#)
[Help](#)

Crawl Status

Crawl Status

Active Crawls: 1
Pending Crawls: 98
Successful Crawls: 3056
Failed Crawls: 121
Crawl Starter: 13m36s

Journal Volume	Crawl Type	Start Time	Duration	Status ¹	Bytes Fetched ²	Pages Fetched ²	Pages Parsed ²	Pages Pending ²	Pages Excluded ²	Not Modified ²	Errors ²	Mime Types ²
Revista Economia e Desenvolvimento Volume 2012 [Universidade Federal da Paraíba]	New	04:22:35	8m6s	Successful	3,653,597	79	37	0	21	0	2	7
Boletim de Geografia Volume 1991	Content	06/03/15										
RACE Volume 2007	New	04:22:29	6306ms	No permission from publisher	0	0	0	0	0	0	1	0
RACE Volume 2007	Content	06/03/15										
RACE Volume 2007	New	04:00:57	21m31s	Successful	13,096,905	149	100	0	141	0	66	7
RACE Volume 2007	Content	06/03/15										
Artcultura Volume 2008	New	03:36:27	24m30s	Successful	25,845,365	220	106	0	18	0	25	7
Artcultura Volume 2008	Content	06/03/15										
ARCHAI: Revista de Estudos sobre as Origens do Pensamento Ocidental Volume 2008	New	03:22:08	14m18s	Successful	4,752,974	139	39	0	18	0	4	7
ARCHAI: Revista de Estudos sobre as Origens do Pensamento Ocidental Volume 2008	Content	06/03/15										
Journal of Animal Behaviour and Biometeorology Volume 2013	New	03:06:05	16m2s	Successful	5,722,760	134	37	0	31	0	4	6
Journal of Animal Behaviour and Biometeorology Volume 2013	Content	06/03/15										
Fórum Linguístico Volume 1999	New	03:05:59	6270ms	No permission from publisher	0	0	0	0	0	0	1	0
Fórum Linguístico Volume 1999	Content	06/03/15										
Revista Ágora Volume 2012	New	02:48:10	17m49s	Successful	4,684,665	127	41	0	26	0	51	7
Revista Ágora Volume 2012	Content	06/03/15										

Figura 21 – Visualização de processo de Crawl

A Figura 20 traz apenas um fragmento de uma extensa listagem que é apresentada nessa funcionalidade, contendo todos os títulos indicados para preservação, de todas as caixas da PLN. A indicação da caixa de origem (no caso, a caixa da UFSC) é possível a partir da seleção do *status* (Figura 22).

Severity ¹	URL	Error
Error	https://periodicos.ufsc.br/index.php/forum/gateway/lockss?year=1999	No permission statement on manifest page.

Notes:

1. Errors and Fatal errors cause the crawl to fail, Warnings do not.

Figura 22 – Visualização – origem da UA

A partir da incorporação das demais sub-redes previstas ou em fase de implantação, como é o caso da sub-rede de teses e dissertações, essa listagem tende a

ficar ainda mais extensa. Segundo o especialista da Aliança LOCKSS, Thibault Yoann Guicherd-Callin, cada nó é livre para determinar a arquitetura da caixa, podendo portanto adotar uma única caixa para preservar todos os tipos de materiais ou caixas individuais, para cada tipo de material ou sub-rede. Essa decisão caberá a cada nó da Rede e não afetará a gestão da mesma.

As instituições parceiras podem então optar por integrar todas as sub-redes da Rede Cariniana ou não, de acordo com interesses e necessidades individuais. Dessa forma, já tendo manifestado interesse em participar da sub-rede de teses e dissertações, prospectando o crescimento da Rede Cariniana, a diversidade de material que se pretende preservar, e a forma como os conteúdos são apresentados nas caixas para monitoramento, sugere-se a adoção de caixas individuais, para cada sub-rede em que a UFRN participar⁹⁵. Com isso, o processo de monitoramento tende a ser mais claro e a identificação de inconsistências de fácil localização.

A implantação de caixas distintas está condicionada a aprovação por parte da equipe de TI da Biblioteca Central Zila Mamede – BCZM, em conjunto com a Direção da Biblioteca e Coordenação do Setor de Repositórios Digitais – SRD. Entretanto, por envolver um setor externo a BCZM, no caso a Superintendência de Informática – SINFO onde o servidor está hospedado, a implantação dependerá ainda de aprovação da Reitoria, para que os recursos necessários sejam adquiridos.

4.2. Boas Práticas

Para além dos requisitos propostos enquanto objetivo principal deste estudo, sugere-se ainda outras recomendações, enquanto boas práticas a serem adotadas pela UFRN, direcionadas aos processos relativos a seu repositório:

⁹⁵ Cada parceiro integral deverá ser consultado quanto a subrede que pretende integrar. O fato dessa parceria ser integral, não condiciona as instituições a participar de todas as subredes previstas pela Cariniana.

4.2.1. Licença de acesso e uso

As teses e dissertações, assim como demais conteúdos informacionais, estão sujeitos a direitos de acesso e uso. Logo, questões relativas aos direitos de autor são determinantes na preservação digital. O RI UFRN oferece aos utilizadores, durante o fluxo de depósito, duas licenças, a Não-exclusiva (Anexo B) e *Creative Commons*. A licença de distribuição não-exclusiva é de uso geral e abrange todos os documentos aceitos pelo repositório, assim como as licenças *Creative Commons*, já adotadas por padrão pelo DSpace.

Com isso, durante o processo de depósito no repositório, são apresentadas aos utilizadores duas licenças, sendo a primeira a Creative Commons, que utilizador poderá recusar, seguida da Não-exclusiva, obrigatória para que os conteúdos sejam disponibilizados. Para o processo de ingestão de conteúdo na Rede Cariniana, o LOCKSS irá coletar o material a partir do reconhecimento de uma licença, que pode ser a Creative Commons ou a não Não-exclusiva. Atualmente existem no RI UFRN materiais com as duas licenças.

Dessa forma, sugere-se utilizar uma única licença padrão para teses e dissertações, a licença Não-exclusiva, que concede à Universidade Federal do Rio Grande do Norte o direito não-exclusivo de reproduzir, converter (como definido abaixo), comunicar e/ou distribuir o documento entregue (incluindo o resumo/abstract) em formato digital ou impresso e em qualquer meio.

4.2.2. Monitoramento da caixa LOCKSS

O monitoramento dos processos desenvolvidos pelo *daemon*, e por consequência, das caixas LOCKSS, está a cargo da equipe da Aliança LOCKSS em conjunto com a coordenação da Rede Cariniana. A instalação da caixa UFRN ocorreu em fevereiro de 2015, e até o momento (junho de 2015) a coordenação da Rede Cariniana detectou uma correção a ser feita, relativa a ativação do manifesto LOCKSS em um título de periódico.

Contudo, dada a natureza dinâmica de publicação das teses e dissertações⁹⁶, em relação à publicação de periódicos, sugere-se que a UFRN adote uma rotina de monitoramento dos processos realizados pelo *daemon*, ao menos uma vez a cada mês, através de sua interface administrativa. Dessa forma, irá identificar e corrigir possíveis inconsistências que impeçam a ingestão de UAs, em tempo hábil, evitando assim que o *daemon* entre em processo de *looping*, na tentativa de reingestão das UAs com algum tipo de problema.

4.2.3. Verificação de vírus

Atualmente nenhum processo de verificação é realizado pela equipe de colaboradores do Setor de Repositórios Digitais – SRD, ao inserir arquivos no RI UFRN por depósito designado⁹⁷. Por esta razão, e levando em consideração o fato de ser o depósito designado o principal meio de alimentação do RI UFRN, sugere-se que seja adotada a verificação dos arquivos antes do depósito no repositório, com o intuito de assegurar que os mesmos estejam livres de vírus.

4.3. Quadro Síntese

Os quadros a seguir apresentam uma síntese do conjunto de requisitos e das boas práticas, respectivamente, a serem recomendados a UFRN no âmbito do acordo de cooperação com a Rede Cariniana.

⁹⁶ Atualmente a UFRN possui cerca de 109 cursos entre mestrados e doutorados, gerando aproximadamente mil e quinhentas teses e dissertações por ano. Fonte: <http://www.ppg.ufrn.br/>

⁹⁷ Quanto a verificação dos arquivos, por parte dos utilizadores que realizam o auto depósito, não é possível fazer nenhum tipo de afirmação neste sentido, uma vez que não há dados ou estudos sobre isto.

Requisitos	
Escopo	Aquilo que se pretende atingir; que materiais se pretende reservar.
Material a ser preservado	Teses e dissertação defendidas na UFRN e em outras instituições.
Stakeholder	Coordenação SRD: responsável por determinar o escopo; Equipe de TI: indicação de parâmetros;
Metadados	Informação estruturada que descreve, explica, localiza, ou de outra forma torna mais fácil a recuperação, uso ou gestão de um recurso de informação.
Padronização	Agregar 19 elementos (Figura 9) às coleções Teses e Dissertações defendidas em outras instituições; Determinar conjunto mínimo de elementos obrigatórios.
Stakeholder	Equipe de TI: configuração das planilhas; Equipe SRD: alimentação dos dados.
Formato	Estrutura que codifica informação em padrões que só podem ser processados e interpretados por combinações específicas de <i>hardware</i> e <i>software</i> .
Adotar formato padrão	PDF / A; Normalização de formatos (para arquivos já existente).
Stakeholder	Direção BCZM e Coordenação SRD: responsável por adotar padrão; Equipe de TI: implantação do PDF/A; normalização do formato.
Nomeação de Arquivos	Conjunto de dados utilizados para identificar os arquivos correspondentes aos itens.
Padronização	Agregar data e truncamento de título.
Stakeholder	Equipe SRD: adoção de novo padrão e correção dos nomes já existentes.
Arquitetura caixa LOCKSS	Definição de utilização das caixas: única para todos os tipos de materiais que se pretende preservar; ou distinta, uma para cada tipo de material ou subrede.
Adoção de caixas distintas	Adotar caixa individual para cada subrede que a instituição participar.
Stakeholder	Direção BCZM e Coordenação SRD: Autorização interna; Reitoria: autorizar aquisição de material; SINFO: implementar.

Tabela 5 – Conjunto de requisitos

Boas Práticas
Licenças
Adotar uma única licença de acesso e uso, para as teses e dissertações.
Monitoramento
Estabelecer rotina para monitorar os processos realizados no daemon, na caixa UFRN; Tempo de verificação: uma vez a cada mês.
Verificação de Vírus
Adotar processo de verificação dos arquivos antes do depósito no repositório, com o intuito de assegurar que os mesmos estejam livres de vírus.

Tabela 6 – Boas práticas

O conjunto de requisitos e boas práticas apresentados, permite tratar a preservação digital a curto-prazo, e deve ser visto como primeiros passos para que a UFRN possa, posteriormente, construir a médio e longo-prazo um plano de preservação digital mais ambicioso, ou mesmo um plano para certificação de seu repositório.

Considerações Finais

A preservação digital depende de um conjunto de atividades complexas, que requer planejamento e tem custos. A adoção de *software* para repositórios digitais que suporta funções voltadas para a preservação, e de tecnologias orientadas à criação de redes colaborativas podem auxiliar significativamente na solução do problema da salvaguarda e acesso de longo prazo dos documentos digitais. A perspectiva metodológica dessas redes colaborativas implica a cooperação entre instituições e exige não só soluções pré-coordenadas em um ambiente tecnológico padronizado, mas também acordos a longo prazo interinstitucionais, que firmem compromissos no intuito de agir em harmonia ao longo do tempo e garantir a preservação, segurança e acesso aos dados.

Nesse sentido a primeira rede colaborativa orientada à preservação de conteúdos digitais foi criada no Brasil pelo IBICT, a Rede Cariniana, com a ambiciosa missão de salvaguardar e possibilitar o acesso a longo prazo dos conteúdos eletrônicos produzidos no país relativos a ciência, tecnologia e patrimônio cultural. Sua criação se deu sob a perspectiva do modelo de preservação digital distribuída, operacionalizada através do *software* LOCKSS, a partir de acordo firmado entre o IBICT e a Aliança LOCKSS.

Com esse acordo, o IBICT e as instituições parceiras da Rede Cariniana, integram uma aliança internacional voltada para a preservação digital de longo prazo. Localmente, a nível de país, além da preservação de longo prazo, a Rede Cariniana provê ainda o compartilhamento de estudos e pesquisas, e a integração de conteúdos da memória institucional digital de forma consorciada e federada.

A UFRN encontrou na Rede Cariniana a possibilidade de salvaguardar seus conteúdos digitais, preservando assim sua memória e garantindo o acesso contínuo a esses conteúdos. Com isso, firmou acordo de cooperação técnica com o IBCT visando a integração da universidade à Rede Cariniana. Essa integração já possibilita a preservação dos Periódicos editados no âmbito da universidade, armazenados e disponibilizados através do SEER, e irá possibilitar a preservação das Teses e Dissertações, armazenadas e disponibilizadas em DSpace pela universidade.

Por ocasião desse acordo de cooperação, enquanto parceira integral e assim como as demais instituições que integram a Rede Cariniana, a universidade concordou em contribuir na especificação de requisitos técnicos, capacitação, criação de serviços, estabelecimento de regras para a produção, e o funcionamento de sistemas de preservação no Brasil. Com o intuito de contribuir para a Rede como um todo, e objetivando especificamente apoiar esse acordo no que se refere a integração DSpace-LOCKSS, este estudo se propôs a desenvolver um conjunto de requisitos mínimos, direcionados para a UFRN e seu repositório, no âmbito dessa cooperação.

Para isso, realizou-se um levantamento bibliográfico que possibilitou identificar os principais conceitos relacionados à área dos repositórios digitais de acesso livre, assim como o *software* utilizado para implantação desses repositórios, o que permitiu compreender tecnicamente como opera o DSpace implantado na UFRN para criação e gerenciamento do RI UFRN. O levantamento possibilitou também identificar conceitos relacionados a área da preservação digital de longo prazo o que, por sua vez, proporcionou a compreensão da perspectiva do modelo de preservação digital distribuída, no qual a Rede Cariniana está fundamentada.

Durante essa discussão teórica, modelos, normas e padrões gerais ou criados para realidades específicas, que fundamentam e fornecem orientação para o funcionamento de sistemas de preservação digital foram apresentados, muito embora não tenham sido discutidos de forma exaustiva. Pontos considerados essenciais foram apresentados, por exemplo, como garantir a integridade e autenticidade dos conteúdos em conformidade com o modelo OAIS e o padrão de metadados Dublin Core, ambos, por sua vez, contemplados no DSpace e no LOCKSS, utilizados pelo RI UFRN e Rede Cariniana, respectivamente.

Além da discussão teórica, realizou-se a análise dos contextos institucionais de cooperação, que permitiu compreender de forma sistêmica as duas estruturas envolvidas neste estudo, o RI UFRN e a Rede Cariniana e ainda: (1) o reconhecimento das relações tecnológicas existentes entre o RI UFRN e a Rede Cariniana; (2) a identificação dos meios que viabilizam a incorporação do acervo digital armazenado no RI UFRN à Rede Cariniana; (3) a identificação dos meios que permitem assegurar a autenticidade e integridade do acervo digital da UFRN; (4) a identificação dos processos utilizados pela Rede Cariniana, que possibilitam a salvaguarda e acesso de

longo prazo. Essa análise dos contextos, em conjunto com a discussão teórica, viabilizaram o desenvolvimento do conjunto de requisitos direcionado à UFRN e seu repositório, no âmbito da cooperação técnica com a Rede Cariniana.

Verificou-se que as relações tecnológicas necessárias entre o RI UFRN e a Rede Cariniana dizem respeito principalmente à instalação da caixa LOCKSS e ao envio de parâmetros para a coordenação da Rede Cariniana, que identificam a caixa da UFRN e permitem a integração na Rede. Para essa instalação, a universidade configurou a sua caixa com especificações superiores às recomendações mínimas exigidas pela Rede e ainda assumiu medidas de segurança, como a virtualização do servidor na SINFO, que por sua vez, mantém política de backup conferindo, assim, maior segurança aos dados.

Quanto a incorporação do acervo digital armazenado no RI UFRN na Rede Cariniana, ficou evidenciado que esta é possível através do *plugin* de integração DSpace-LOCKSS. Apesar do primeiro teste realizado no IBICT não ter obtido êxito, no que se refere a coleta das teses e dissertações para ingestão no LOCKSS, outros testes realizados após a correção dos parâmetros do *plugin*, com um repositório da *University of Virginia*, lograram êxito. Com isso, se pode afirmar que será possível iniciar os processos propriamente ditos, relativos às teses e dissertações. No entanto, as especificações técnicas com as alterações necessárias para configuração do *plugin*, foram enviadas à UFRN em 16 de junho de 2015, não havendo portanto tempo hábil para realizar e discutir os novos testes que serão realizados, já no âmbito da UFRN e com o acervo digital do RI UFRN e não mais com um repositório clone.

Referente a garantia de autenticidade e integridade do acervo digital da UFRN, mostrou-se que é possível através de ações realizadas durante o armazenamento dos conteúdos na caixa LOCKSS UFRN, que garante a preservação a nível de *bit*, como também pelo próprio processo de preservação digital distribuída via P2P, que garante a salvaguarda dos conteúdos originais de forma autêntica, e promove a recuperação de conteúdos que venham a ser corrompidos. E, através de qualquer uma das caixas que forma a PLN IBICT, é possível a recuperação e o acesso aos conteúdos salvaguardados.

Assim, a discussão teórica especialmente no que se referiu aos modelos, normas e padrões que orientam processos e sistemas para a preservação digital,

viabilizou os requisitos propostos. O conjunto prevê requisitos mínimos direcionados principalmente para a preparação dos conteúdos, antes da ingestão e armazenamento. Essa necessidade foi evidenciada por meio do conhecimento empírico adquirido com a subrede de periódicos, que mostrou a importância de trabalhar os conteúdos para ingestão. Dessa forma, foram propostos os requisitos relativos a: escopo; metadados; formato; nomeação URI; nomeação de arquivos; e arquitetura caixa LOCKSS.

O requisito escopo foi sugerido para que os materiais a serem preservados estejam de acordo com a missão do RI UFRN. Os requisitos metadados, formato, e nomeação de arquivos foram sugeridos visando a preparação e padronização dos conteúdos, com fins de ingestão e o armazenamento. Quanto à arquitetura da caixa, sugeriu-se que a UFRN adote caixas individuais para cada subrede da qual fará parte. Acredita-se que com isso, será possível monitorar os conteúdos armazenados e processos realizados em cada caixa LOCKSS.

Além desses requisitos, e buscado suprir lacunas identificadas no contexto do RI UFRN, foi proposta a adoção de uma rotina de monitoramento da caixa LOCKSS, para que não fique apenas a cargo da coordenação da Rede Cariniana e de Stanford, e ainda para corrigir possíveis inconsistências em tempo hábil. Foi proposta também, a adoção da licença de distribuição não-exclusiva como única licença de acesso e uso utilizada para o depósito e disponibilização das teses e dissertação. E por fim, adoção de processo de verificação de vírus, para assegurar que os arquivos a serem ingeridos estarão livres de vírus.

No entanto, para o desenvolvimento deste estudo algumas dificuldades foram encontradas. O primeiro teste de integração DSpace - LOCKSS realizado entre os dias 23 e 25 de março de 2015 em Brasília, não apresentou o resultado esperado, uma vez que a coleta das teses e dissertações não foi possível por uma inconsistência no *plugin*. Com isso, a análise desse processo de coleta e a identificação das UAs e consequente verificação dos conteúdos na caixa LOCKSS, não foi possível. Essa análise era considerada essencial, pois a partir dela outras necessidades poderiam ter sido identificadas e como consequência, outros requisitos poderiam ter sido propostos.

Muito embora novos testes tenham sido realizados com êxito pela equipe da Aliança LOCKSS, utilizando um repositório da *University of Virginia*, os resultados obtidos com esse teste não foram divulgados até a finalização deste estudo, não sendo possível portanto perceber como será o armazenamento, no que se refere a identificação e visualização dos conteúdos nas caixas.

Novas especificações técnicas para a instalação do *plugin* de integração LOCKSS-DSpace, a partir do êxito alcançado nos testes realizados com repositório da *University of Virginia*, foram enviadas ao IBICT, que por sua vez as repassou à UFRN em 16 de junho de 2015. Novos testes com base nessas especificações serão realizados, dessa vez no âmbito da UFRN com as teses e dissertações que encontram-se no RI UFRN. Contudo, não será possível discuti-los neste estudo, tão pouco analisá-los, por não haver tempo hábil.

Outra dificuldade que cabe mencionar é o fato de haver pouco material acerca dos mecanismos e processos empregados pelo LOCKSS, a informação apresentada neste estudo se embasou: no livro **A Guide to distributed digital preservation**, utilizado também pela equipe do IBICT como referencial para o desenvolvimento dos primeiros documentos normativos da Rede Cariniana; em informação disponibilizada no *site* do LOCKSS, considerada geralista, portanto com pouco aprofundamento; e ainda, na experiência empírica adquirida pela equipe IBICT a partir do uso do *software* e da interação com a equipe da Aliança LOCKSS, disponibilizada através do site da Rede Cariniana na forma de slides, utilizados para apresentações diversas da Rede Cariniana e ainda para realização de treinamentos. Segundo o especialista Thibault Yoann Guicherd-Callin, que esteve no IBICT para a realização dos primeiros testes de integração, essa é uma falha da equipe que desenvolveu e trabalha com o *software*, e que só recentemente iniciou o desenvolvimento de documentação sobre o LOCKSS e seu funcionamento.

Embora algumas dificuldades tenham sido encontradas, acredita-se que o objetivo desse estudo foi alcançado, e que com o conjunto de requisitos e boas praticas propostos, será possível apoiar a cooperação entre a UFRN e a Rede Cariniana a curto prazo, e com isso, permitir que a UFRN construa a médio e longo prazo um plano de preservação mais ambicioso, ou mesmo um programa para certificação de seus repositórios.

Quanto ao conjunto de requisitos apresentados, perspectiva-se a aplicação do mesmo, em sua integralidade, no âmbito da UFRN, e espera-se que possa ser utilizado pelas demais instituições participantes da Rede Cariniana, a fim de adequar e padronizar os conteúdos a serem preservado, bem como o ambiente da cooperação.

Apesar de reconhecer que o objetivo proposto neste estudo foi alcançado, cabe ressaltar que a realização de novos testes de integração LOCKSS-DSpace é essencial para analisar o comportamento dos *software*, o que poderá apontar novas necessidades não identificadas neste estudo. Reconhece-se ainda a necessidade de avaliar a aplicação do conjunto de requisitos apresentados, e com isso, verificar a eficácia dos requisitos para o processo de preservação e para a cooperação com a Rede Cariniana.

Referências Bibliográficas

ANDRADE, Morgana Carneiro et al. Esquemas de metadados utilizados por repositórios digitais científicos e de objetos educacionais: estudo preliminar. **Cadernos BAD**, n. 2, p. 219-222, jul-dez 2014.

ASSOCIAÇÃO PARA PROMOÇÃO E DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO. **Glossário da Sociedade de Informação**. 2011. Disponível em: <<http://www.apdsi.pt/uploads/news/id432/gloss%C3%A1rio%20da%20si%20-%20vers%C3%A3o%202011.pdf>>. Acesso em: 03 dez.2014.

AUTHENTICITY TASK FORCE. **Conceptual Framework for the Requirements for Assessing and Maintaining the Authenticity of Electronic Records**. 2002. Disponível em: <http://www.interpares.org/book/interpares_book_k_appo2.pdf>. Acesso em: 04 dez.2014.

BROWN, Adrian. **Digital Preservation Guidance Note 1: Selecting file formats for long-term preservation**. 2008. Disponível em: <<https://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>>. Acesso em: jun. 2015.

CABRAL, Luiz Henrique. **Integração RI UFRN e Rede Cariniana**: depoimento [abr. 2015]. Entrevistador: Elisângela Alves de Moura. Entrevista concedida no âmbito da dissertação Repositórios e preservação digital: proposta de requisitos para a integração do RI UFRN com a Rede Cariniana.

CAÇÃO, Rosário. **Análise Qualitativa de Dados com o NVIVO**. 2009. Disponível em: <<http://pt.slideshare.net/rosariocacao/anlise-qualitativa-de-dados-com-o-nvivo-1334976>>. Acesso em: 12 jan.2015.

CAPLAN, Priscilla. **Understanding PREMIS**. 2009. Disponível em: <<http://www.loc.gov/standards/premis/understanding-premis.pdf>>. Acesso em: 04 dez.2014.

CARINIANA. Rede Brasileira de Serviços de Preservação Digital. **Reunião técnica**. 2014. Disponível em: <<http://pt.slideshare.net/redecariniana/relatrio-cariniana-2014>>. Acesso em: 11 maio 2015.

_____. **Plano de trabalho das equipes da rede brasileira de serviços de preservação digital cariniana**: projeto de implementação e integração da rede brasileira de serviços de preservação digital – cariniana (2013-2018). Brasília: IBICT, 2013. Disponível em: <<http://cariniana.ibict.br/index.php/publicacoes/item/18-plano-de-trabalho-cariniana-2014-2018>>. Acesso em: 02 maio 2015.

CASTELLS, Manuel. **A galáxia internet**: reflexões sobre a internet, negócios e sociedade. Lisboa: Fundação Calouste Gulbenkian, 2004.

CASTRO, Cristiane Yanase Hirabara de et al. Repositórios institucionais confiáveis: repositório institucional como ferramenta para a preservação digital. In: SAYÃO, Luis Fernando et al. **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p.283-304.

Disponível em:

<https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.pdf>. Acesso em: 04 dez.2014.

CROW, Raym. The Case for Institutional Repositories: a SPARC position paper. **ARL Bimonthly Report** **223**. 2002. Disponível em:

<http://www.sparc.arl.org/sites/default/files/media_files/instrepo.pdf>. Acesso em: 04 dez.2014.

DIGGORY, Mark; LUYTEN, Bram. **Functional Overview**: DSpace 4.x Documentation. 2013. Disponível em:

<<https://wiki.duraspace.org/display/DSDOC4x/Functional+Overview>>. Acesso em: dez. 2014.

DIGITAL PRESERVATION COALITION. **The Preservation Management of Digital Material Handbook**. 2008. Disponível em:

<<http://www.dpconline.org/pages/handbook/docs/DPCHandbook.pdf>>. Acesso em: 04 dez.2014.

DODEBEI, Vera. Repositórios institucionais: por uma memória criativa no ciberespaço. In: SAYÃO, Luis Fernando et al. In: **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p.83-106. Disponível em:

<https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.pdf>. Acesso em: 04 dez. 2014.

DSpace. **About DSpace**. Disponível em: <<http://www.dspace.org/introducing>>. Acesso em: dez. 2014.

ELSTROM, Gry Vindelev; SIERMAN, Barbara. **Guidance policy elements and preservation procedure elements**. 2014. Disponível em: <<http://wiki.opf-labs.org/display/SP/Policy+Elements>>. Acesso em: jun. 2015.

EMC (2014). **Universo Digital é invadido por sensores**. Press Release. 2014. Disponível em: <<http://portugal.emc.com/about/news/press/2014/20140409-01.htm>>. Acesso em 30 dez. 2014.

FERREIRA, Carla Alexandra Silva. **Preservação da informação Digital**: uma perspectiva orientada para as bibliotecas. Dissertação (Mestrado em Informação, Comunicação e Novos Media) – Faculdade de Letras da Universidade de Coimbra, Coimbra, 2011. Disponível em:

<<https://estudogeral.sib.uc.pt/bitstream/10316/15001/1/Preservacao%20da%20Informacao%20Digital.pdf>>. Acesso em: 04 dez.2014.

FERREIRA, José Miguel Araújo. **Preservação de longa duração de informação digital no contexto de um arquivo histórico**. Tese (Doutorado em Tecnologias

e Sistemas de Informação, na especialidade de Sociedade da Informação) - Escola de Engenharia da Universidade do Minho, Guimarães, 2009. Disponível em: <<http://repositorium.sdum.uminho.pt/handle/1822/9563>>. Acesso em: 04 dez.2014.

FERREIRA, Miguel. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos. Guimarães: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <<http://repositorium.sdum.uminho.pt/handle/1822/5167>>. Acesso em: 04 dez.2014.

FERREIRA, Miguel; BAPTISTA, Ana Alice; RAMALHO, José Carlos. **Avaliação automática de redes distribuídas de conversores**. 2005. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/3616/1/CAPSI05-vo.10.pdf>>. Acesso em: 22 dez.2014.

FERRER, Igor Dias. **Rede Cariniana: componentes tecnológicos**: depoimento [maio 2015]. Entrevistador: Elisângela Alves de Moura. Entrevista concedida no âmbito da dissertação Repositórios e preservação digital: proposta de requisitos para a integração do RI UFRN com a Rede Cariniana.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009. Disponível em: <<http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>>. Acesso em: 15 jan.2015.

GONZÁLEZ REY, Fernando Luis. **Pesquisa qualitativa e subjetividade**: os processos de construção da informação. São Paulo: Pioneira Thomson Learning, 2005.

GUEDES, Clediane de Araújo. **Setor de Repositórios Digitais da Biblioteca Central Zila Mamede**: depoimento [abr. 2015]. Entrevistador: Elisângela Alves de Moura. Entrevista concedida no âmbito da dissertação Repositórios e preservação digital: proposta de requisitos para a integração do RI UFRN com a Rede Cariniana.

HEDSTROM, Margaret. Digital Preservation: problems and prospects. **Digital Libraries**, n.20, Mar. 2001. Disponível em: <http://www.dl.slis.tsukuba.ac.jp/DLjournal/No_20/1-hedstrom/1-hedstrom.html>. Acesso em: 23 dez.2014.

HOFMAN, Hans. **Can Bits and Bytes be Authentic?** Preserving the Authenticity of Digital Objects. 2002. Disponível em: <http://eprints.erpanet.org/39/01/hofman_glasgow02.pdf>. Acesso em: 16 dez.2014.

IBICT. **Guia de preservação digital de periódicos eletrônicos na Rede Cariniana**. Brasília: IBICT, 2015.

IBICT. **Rede Cariniana**. Disponível em: <<http://cariniana.ibict.br/>>. Acesso em: jan.2015.

INTERPARES. **Terminology Database**. [Em linha]. [Consult.03 jan. 2015].

Disponível em:

http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=a&term=70

KUNY, Terry. **A Digital Dark Ages?** challenges in the preservation of electronic Information. 1997. Disponível em: <<http://archive.ifla.org/IV/ifla63/63kuny1.pdf> >. Acesso em: 04 dez. 2015.

LEE, Kyong-Ho et al. The State of the Art and Practice in Digital Preservation. **J. Res. Natl. Inst. Stand. Technol.** v. 107, n.1, p.93–106, 2002. Disponível em: <https://ia600705.us.archive.org/23/items/jresv107n1p93/jresv107n1p93_A1b.pdf> Acesso em: 04 dez. 2015.

LEITE, Fernando Cesar Lima. **Como gerenciar e ampliar a visibilidade da informação científica brasileira:** repositórios institucionais de acesso aberto. Brasília: IBICT, 2009.

LEITE, Fernando César Lima; COSTA, Sely. Repositórios institucionais como ferramentas de gestão do conhecimento científico no ambiente acadêmico. **Perspect. Ciênc. Inf.**, Belo Horizonte, v.11 n.2, p. 206 -219, mai./ago. 2006.

LOCKSS. **About LOCKSS**. Disponível em: <<http://www.lockss.org/about/how-it-works/#sustainable-format-migration>>. Acesso em: abr. 2015.

LYNCH, Clifford A. Institutional Repositories: essential infrastructure for scholarship in the digital age. **ARL: A Bimonthly Report**, n. 226, feb. 2003. Disponível em: <http://scholarship.utm.edu/21/1/Lynch_IRs.pdf>. Acesso e,: 02 jan. 2015.

MACNEIL, Heather et al. **Authenticity Task Force Report**. 2002. Disponível em: <http://www.interpares.org/display_file.cfm?doc=ip1_atf_report.pdf>. Acesso em: 04 dez. 2015.

MARCONDES, Carlos H. (org.). **Bibliotecas digitais:** saberes e práticas. Brasília: IBCT, 2005.

MÁRDERO ARELLANO, Miguel Ángel. Preservação de documentos digitais. **Ci. Inf.**, Brasília, v. 33, n. 2, p. 15-27, maio/ago. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a02v33n2.pdf>>. Acesso em: 04 dez. 2015.

_____. **Critérios para a preservação digital da informação científica**. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, 2008. Disponível em: <http://bdtd.bce.UnB.br/tesesimplificado/tde_busca/arquivo.php?codArquivo=4547>. Acesso em: 04 dez. 2015.

_____. Cariniana: uma rede nacional de preservação digital. **Ci. Inf.**, Brasília, v. 41 n. 1, p.83-91, jan./abr. 2012. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2127/1794>>. Aceso em: 10 jan. 2015].

_____. **Rede Carinaina**: depoimento [maio 2015]. Entrevistador: Elisângela Alves de Moura. Entrevista concedida no âmbito da dissertação Repositórios e preservação digital: proposta de requisitos para a integração do RI UFRN com a Rede Cariniana.

MARTINS, Lourdes. **Entrevista semi-estruturada**. 2010. Disponível em: <<http://pt.slideshare.net/mlurdesmartins/entrevista-semi-estruturada>>. Acesso em: 12 jan. 2015.

MORENO, Fernanda Passini; LEITE, Fernando César Lima; MÁRDERO ARELLANO, Miguel Ángel. Acesso livre a publicações e repositórios digitais em ciência da informação no Brasil. **Perspect. ciênc. inf.**, Belo Horizonte, v.11, n.1, p. 82-94, jan./abr. 2006. Disponível em: <<http://www.scielo.br/pdf/pci/v11n1/v11n1a07.pdf>>. Acesso em: 03 dez. 2014.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Understanding Metadata**. Bethesda: NISO, 2004. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. Acesso em: 02 jan. 2015.

NATIONAL LIBRARY OF AUSTRALIA. **Directrices Para La Preservación Del Patrimonio Digital**. UNESCO, 2003. Disponível em: <<http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>>. Acesso em: 04 dez. 2015.

OCLC. **Preservation Metadata and the OAIS Information Model**: a metadata Framework to Support the Preservation of Digital Objects: A Report by The OCLC/RLG Working Group on Preservation Metadata. 2002. Disponível em: <http://www.oclc.org/content/dam/research/activities/pmwg/pm_framework.pdf?urlm=161391>. Acesso em: 03 dez. 2014.

OCLC/RLG Working Group on Digital Archive Attributes. **Trusted digital repositories**: attributes and responsibilities. 2002. Disponível em: <<http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf?urlm=161690>>. Acesso em: 16 maio 2015.

OCLC/RLG Working Group on Preservation Metadata. **Preservation Metadata for Digital Objects**: a review of the state of the art. 2001. Disponível em: <http://www.oclc.org/content/dam/research/activities/pmwg/presmeta_wp.pdf?urlm=161396>. Acesso em: 16 dez. 2014.

OETTLER, Alexandra. **PDF/A in a Nutshell 2.0**: PDF for long-term archiving. 2013. Disponível em: <http://www.pdfa.org/wp-content/uploads/2013/05/PDFA_in_a_Nutshell_211.pdf>. Acesso em: jun. 2015.

PACHECO, Gonzalo Lara et al. **Digitalización de colecciones**: texto e imagen. México: Universidad Nacional Autónoma de México, 2008. v.1. Disponível em: <<http://www.digitalizacion.unam.mx/>>. Acesso em: 02 jan. 2015.

OLIVEIRA, Alexandre Faria de. **Manual de instalação LOCKSS**. Brasília: IBICT, 2013.

PDF/A – Archiving. **Information about PDF/A**. Disponível em: <<http://www.pdf-tools.com/pdf/pdfa-longterm-archiving-iso-19005-pdf.aspx#PDFAinfos>>. Acesso em: jun. 2015.

PINTO, Maria Manuela. **PRESERVMAP**: um roteiro da preservação na era digital. Porto: Edições Afrontamento, 2009.

PREMIS Editorial Committee. **PREMIS Data Dictionary for Preservation Metadata version 2.0**. 2008. Disponível em: <<http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>>. Acesso em: 03 dez. 2015.

RODRIGUES, Eloy. Acesso livre ao conhecimento: a mudança do sistema de comunicação da ciência e os profissionais de informação. **Cadernos BAD**, v.1, 2004. Disponível em: <<http://repositorium.sdum.uminho.pt/handle/1822/670>>. Acesso em: 03 dez. 2015.

ROG, Judith; VAN WIJK, Caroline. **Evaluating File Formats for Long-term Preservation**. Disponível em: <https://www.kb.nl/sites/default/files/docs/KB_file_format_evaluation_method_27022008.pdf>. Acesso em jun. 2015.

ROSENTHAL, David S. H. et al. Requirements for Digital Preservation Systems: a bottom-up approach. **D-Lib Magazine**, v.11, n. 11, nov. 2005. Disponível em: <<http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>>. Acesso em: fev. 2015.

ROTHENBERG, J. **Avoiding technological quicksand**: finding a viable technical foundation for digital preservation: a report to the Council on Library and Information Resources. Washington: Council on Library and Information Resources, 1999. Disponível em: <<http://www.clir.org/pubs/reports/rothenberg/pub77.pdf>>. Acesso em: 12 jan. 2015.

RUUSALEPP, Raivo et al. Standards alignment. p.115-165 In: MCGOVERN, Nancy Y.; SKINNER, Katherine. **Aligning national approaches to digital preservation**. Atlanta: Educopia institute publications, 2012. Disponível em: <http://educopia.org/sites/educopia.org/files/publications/Aligning_National_Approaches_to_Digital_Preservation.pdf>. Acesso em: jun. 2015.

SAGIROGLU, Seref; SINANC, Duygu. **Big Data**: a review. 2013. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=06567202>>. Acesso em: 02 jan. 2015.

SARAMAGO, Maria de Lurdes. **Metadados para preservação digital e aplicação do modelo OAIS**. 2004. Disponível em: <<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/640>>. Acesso em: 16 dez. 2014.

SAYÃO, Luis Fernando. **Identificadores Persistentes**: no contexto das bibliotecas digitais. Disponível em: <<http://www.stf.jus.br/arquivo/sijed/14.pdf>>. Acesso em: jun. 2015.

SAYÃO, Luis Fernando; MARCONDES, Carlos Henrique. Software livres para repositórios institucionais: alguns subsídios para a selecção. In: SAYÃO, Luis Fernando et al. **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p.9-22. Disponível em: <https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.pdf>. Acesso em: 04 dez. 2014.

SHINTAKU, Milton; MEIRELLES, Rodrigo. **Manual do DSpace**: administração de repositórios. Salvador: EDUFBA, 2010.

SKINNER, Katherine; SCHULTZ, Matt. **A Guide to distributed digital preservation**. Atlanta, Georgia: Educopia Institute, 2010. Disponível em: <http://www.metaarchive.org/sites/metaarchive.org/files/GDDP_Educopia.pdf>. Acesso em: 10 jan.2015.

SOUZA, Arthur Heleno Lima R. de. **Guia da Interface Administrativa LOCKSS**. 2013. Disponível em: <<http://cariniana.ibict.br/index.php/publicacoes/item/1-manual-de-ial>>. Acesso em: abr. 2015.

_____. **Curso de preservação digital**: sistema LOCKSS. 2014. Disponível em: <http://pt.slideshare.net/redecariniana/curso-lockss?qid=45df0bfa-0534-4b2a-924c-5ffcf620ee49&v=default&b=&from_search=2>. Acesso em: 02 maio 2015.

_____. **Relatório Técnico**: análise das caixas LOCKSS. Brasília: IBICT, 2014b. Disponível em: <<http://cariniana.ibict.br/index.php/publicacoes/item/28-relatorio-tecnico-caixas-lockss>>. Acesso em: abr. 2015.

SOUZA, Arthur Heleno Lima Rodrigues de et al. **O modelo de referência OAIS e a preservação digital distribuída**. Ci. Inf., Brasília, DF, v. 41 n. 1, p.65-73, jan./abr. 2012. Disponível em: <<http://revista.ibict.br/cienciadainformacao/index.php/ciinf/article/view/2118/1792>>. Acesso em: 10 jan. 2015.

STANFORD UNIVERSITY LIBRARIES. **Web Archiving**. Disponível em: <<https://library.stanford.edu/projects/web-archiving>>. Acesso em: jun. 2015.

TASK FORCE. **Preserving digital information**: report of the Task Force on Archiving of Digital Information. Washington, 1996. Disponível em: <<http://www.clir.org/pubs/reports/pub63watersgarrett.pdf>>. Aceso em: 05 jan. 2015.

THIBODEAU, Kenneth. **Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years**. 2002. Disponível em: <<http://www.clir.org/pubs/reports/pub107/thibodeau.html>>. Acesso em: 04 dez. 2014.

THIBODEAU, Kenneth et al. **Preservation Task Force Report**. Disponível em: <http://www.interpares.org/book/interpares_book_f_part3.pdf>. Acesso em: 04 maio 2015.

THOMAS, Chuck; MCDONALD, Robert H. Measuring and Comparing Participation Patterns In Digital Repositories, Repositories by the Numbers Part I. **D-Lib Magazine**, v. 13, n. 9/10, 2007. Disponível em: <<http://www.dlib.org/dlib/september07/mcdonald/o9mcdonald.html>>. Acesso em: 23 dez. 2014.

UNESCO. **Charter on the Preservation of Digital Heritage**. 2003. Disponível em: <http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html>. Acesso em: 04 dez. 2014.

ANEXOS

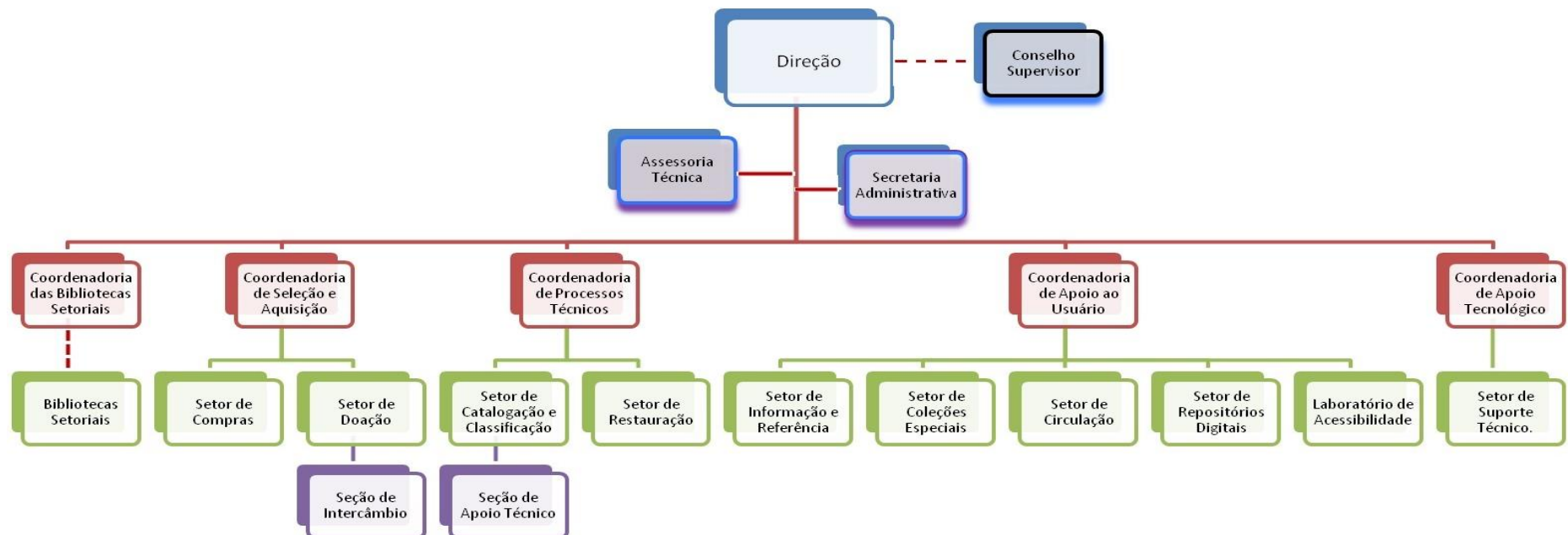
Anexo A – Organograma BCZM



Ministério da Educação
Universidade Federal do Rio grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Biblioteca Central Zila Mamede – BCZM



Organograma da BCZM



Anexo B – Licença de distribuição não-exclusiva

LICENÇA DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

Ao assinar e entregar esta licença, o/a Sr./Sra. (autor ou detentor dos direitos de autor):

a) Concede à Universidade Federal do Rio Grande do Norte o direito não-exclusivo de reproduzir, converter (como definido abaixo), comunicar e/ou distribuir o documento entregue (incluindo o resumo/abstract) em formato digital ou impresso e em qualquer meio.

b) Declara que o documento entregue é seu trabalho original, e que detém o direito de conceder os direitos contidos nesta licença. Declara também que a entrega do documento não infringe, tanto quanto lhe é possível saber, os direitos de qualquer outra pessoa ou entidade.

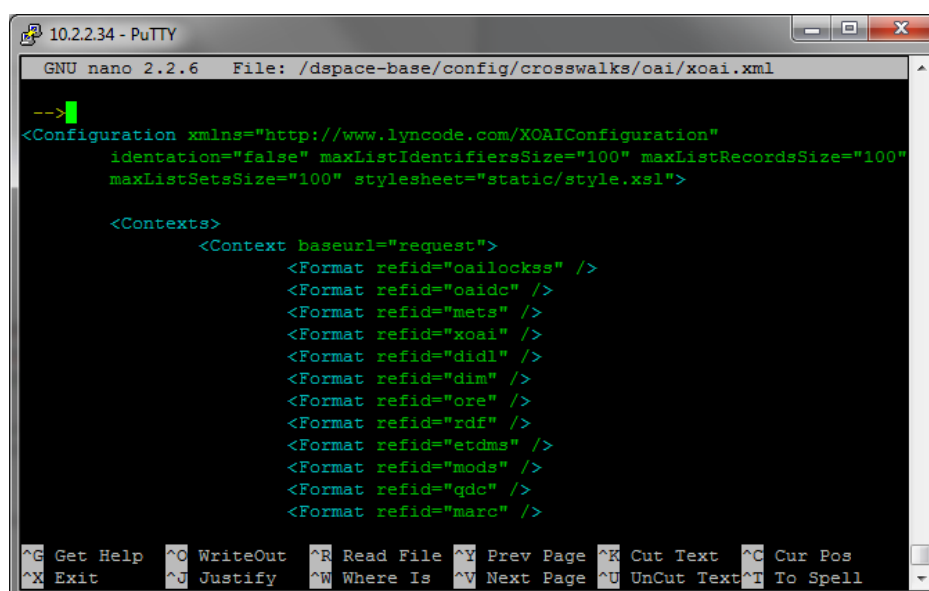
c) Se o documento entregue contém material do qual não detém os direitos de autor, declara que obteve autorização do detentor dos direitos de autor para conceder à Universidade Federal do Rio Grande do Norte os direitos requeridos por esta licença, e que esse material cujos direitos são de terceiros está claramente identificado e reconhecido no texto ou conteúdo do documento entregue. Se o documento entregue é baseado em trabalho financiado ou apoiado por outra instituição que não a Universidade Federal do Rio Grande do Norte, declara que cumpriu quaisquer obrigações exigidas pelo respectivo contrato ou acordo.

A Universidade Federal do Rio Grande do Norte identificará claramente o(s) seu(s) nome(s) como o(s) autor(es) ou detentor(es) dos direitos do documento entregue, e não fará qualquer alteração, para além das permitidas por esta licença.

Anexo C – Configuração do plugin de integração LOCKSS - DSpace⁹⁸

Para a instalação do plugin de interoperabilidade LOCKSS/DSpace, foram realizados os seguintes procedimentos:

1. Conforme consultado no passo a passo, primeiro editou-se o arquivo “[dspace]/config/crosswalks/oai/xoai.xml” e:
 - a. Abaixo da tag “<Context baseurl=“request”>” foi adicionado o campo: “<Format refid=“oailockss” />”



- b. Abaixo da tag “<Formats>” adicionou-se o seguinte fragmento:

```
<!-- Format to provide LOCKSS permission statements -->
<Format id="oailockss">
  <Prefix>oai_lockss</Prefix>
  <XSLT>metadataFormats/oai_lockss.xsl</XSLT>
  <Namespace>http://www.openarchives.org/OAI/2.0/oai\_dc/</Namespace>
  <SchemaLocation>http://www.openarchives.org/OAI/2.0/oai\_dc.xsd</SchemaLocation>
</Format>
```

⁹⁸ Diretrizes fornecidas pela equipe técnica da UFRN

```

10.2.2.34 - PuTTY
GNU nano 2.2.6 File: /dspace-base/config/crosswalks/oai/xoai.xml

        <Format refid="oaidc" />
        <Format refid="mets" />
    </Context>
</Contexts>

<Formats>
    <!-- Format to provide LOCKSS permission statements -->
    <Format id="oailockss">
        <Prefix>oai_lockss</Prefix>
        <XSLT>metadataFormats/oai_lockss.xsl</XSLT>
        <Namespace>http://www.openarchives.org/OAI/2.0/oai_dc/<$
        <SchemaLocation>http://www.openarchives.org/OAI/2.0/oai$
    </Format>
    <Format id="oaidc">
        <Prefix>oai_dc</Prefix>
        <XSLT>metadataFormats/oai_dc.xsl</XSLT>
        <Namespace>http://www.openarchives.org/OAI/2.0/oai_dc/<$
        <SchemaLocation>http://www.openarchives.org/OAI/2.0/oai$
    </Format>

```

2. Realizou-se a cópia do arquivo

“[dspace]/config/crosswalks/oai/metadataFormats/oai_dc.xsl”:

- cp /dspace-base/config/crosswalks/oai/metadataFormats/oai_dc.xsl
/dspace-base/config/crosswalks/oai/metadataFormats/oai_lockss.xsl

3. Em seguida editou-se o novo arquivo **“oai_lockss.xsl”** e, abaixo da tag **“<oai_dc:dc>”** foi adicionado o seguinte fragmento:

```

<!-- LOCKSS permission statement -->
<dc:rights>LOCKSS system has permission to collect, preserve, and serve
this Archival Unit</dc:rights>

```

```

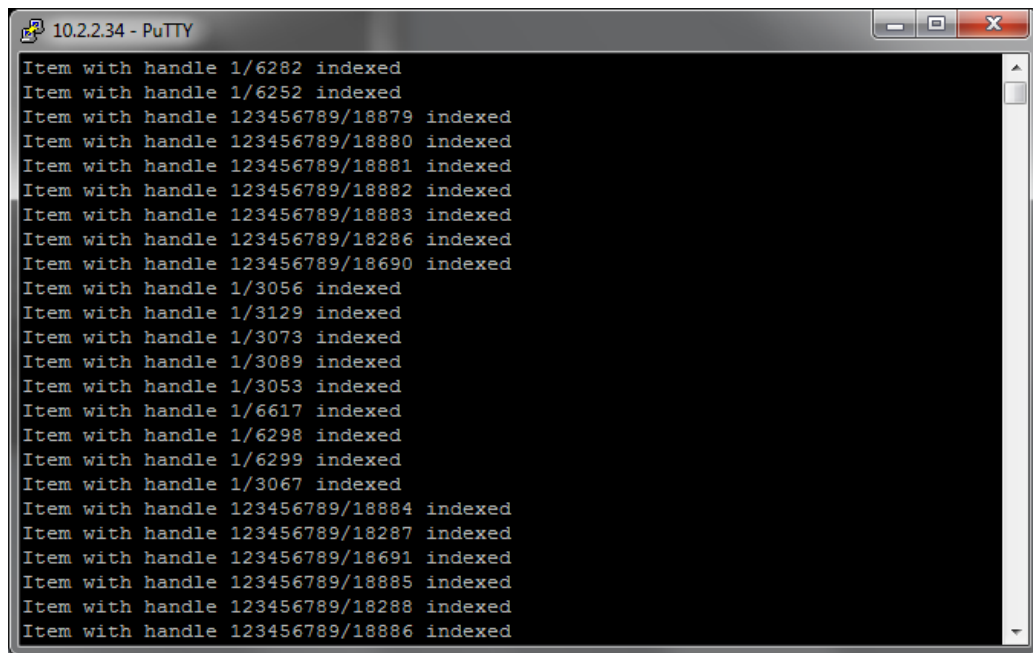
10.2.2.34 - PuTTY
GNU nano 2.2.6 File: /dspace-base/config/crosswalks/oai/metadataFormats/oai_lockss.xsl

-->
<xsl:stylesheet
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:doc="http://www.lyncode.com/xoai"
  version="1.0">
  <xsl:output omit-xml-declaration="yes" method="xml" indent="yes" />
  <xsl:template match="/">
    <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
      <!-- LOCKSS permission statement -->
      <dc:rights>LOCKSS system has permission to collect, preserve, and serve
      this Archival Unit</dc:rights>

      <!-- dc.title -->
    </oai_dc:dc>
  </template>
  <xsl:for-each select="doc:metadata/doc:element[@name='dc']/doc:element[@name='title']/doc:element">

```

- Foi executado o comando **“/dspace-base/bin/dspace oai clean-cache”**
- Excepcionalmente, por não ter sido executado em nenhum momento desde a instalação do repositório clone, foi executado o seguinte comando:
“dspace oai import -c -v”



```
10.2.2.34 - PuTTY
Item with handle 1/6282 indexed
Item with handle 1/6252 indexed
Item with handle 123456789/18879 indexed
Item with handle 123456789/18880 indexed
Item with handle 123456789/18881 indexed
Item with handle 123456789/18882 indexed
Item with handle 123456789/18883 indexed
Item with handle 123456789/18286 indexed
Item with handle 123456789/18690 indexed
Item with handle 1/3056 indexed
Item with handle 1/3129 indexed
Item with handle 1/3073 indexed
Item with handle 1/3089 indexed
Item with handle 1/3053 indexed
Item with handle 1/6617 indexed
Item with handle 1/6298 indexed
Item with handle 1/6299 indexed
Item with handle 1/3067 indexed
Item with handle 123456789/18884 indexed
Item with handle 123456789/18287 indexed
Item with handle 123456789/18691 indexed
Item with handle 123456789/18885 indexed
Item with handle 123456789/18288 indexed
Item with handle 123456789/18886 indexed
```

6. Restartou-se o tomcat (servlet container) e foi recebido o feedback positivo da equipe do IBICT.

APÊNDICES

APÊNDICE A – GUIÕES DE ENTREVISTA



INVESTIGAÇÃO DE MESTRADO

Repositórios e preservação digital: proposta de requisitos para a integração do RI UFRN com a Rede Cariniana.

Mestranda | Elisângela Alves de Moura

Orientadora | Doutora Maria Cristina de Carvalho Alves Ribeiro

Coorientador | Doutor Miguel Ángel Márdero Arellano

GUIÃO DE ENTREVISTA

NOTA INTRODUTÓRIA:

- Apresentação individual
- Apresentação do estudo

À coordenação do Setor de Repositórios Digitais (SRD) – Biblioteca Central Zila Mamede (BCZM) / RI UFRN

1. De acordo com a política de informação, o RI UFRN é gerido por uma comissão. Como essa comissão atua na gestão do repositório?

a) Em que momento ou sob quais circunstâncias a gestão do Ri UFRN passou a ser de responsabilidade da BCZM?

b) Há formalização quanto ao papel da BCZM na gestão do repositório?

2. Ainda de acordo com a política de informação, o depósito de itens no RI UFRN se dá prioritariamente através do auto arquivamento. É mesmo o auto arquivamento a principal forma de submissão?

a) É possível mensurar isso?

b) As teses e dissertações, especificamente, são depositadas no repositório por auto depósito ou depósito designado?

c) E como é realizado o depósito designado?

3. No âmbito da cooperação com a Rede Cariniana, a UFRN elaborou plano de trabalho para integração DSpace – LOCKSS?

a) Alguma ação/estratégia foi empreendida neste sentido?

4. Com relação a ingestão de conteúdos no DSpace, há um padrão quanto ao conjunto de metadados utilizados para as teses e dissertações?

5. Quanto aos arquivos, há padrão no que se refere a nomeação, tamanho ou formato desses arquivos?

a) E quanto ao formato dos arquivos?

6. Por padrão, o DSpace utiliza o Handle System para a identificação persistente dos objetos. A universidade realizou registro junto ao Handle.Net Registry Service – HNRS?

a) A instituição pretende realizar o registro?

INVESTIGAÇÃO DE MESTRADO

Repositórios e preservação digital: proposta de requisitos para a integração do RI UFRN com a Rede Cariniana.

Mestranda | Elisângela Alves de Moura

Orientadora | Doutora Maria Cristina de Carvalho Alves Ribeiro

Coorientador | Doutor Miguel Ángel Márdero Arellano

GUIÃO DE ENTREVISTA

NOTA INTRODUTÓRIA:

- Apresentação individual e institucional

- Apresentação do estudo

Ao Analista de sistemas da Biblioteca Central Zila Mamede (BCZM) / UFRN

1. A Rede Cariniana fez recomendações expressas quanto as especificações mínimas para instalação da caixa LOCKSS UFRN. Essas recomendações foram seguidas?

a) Quais as configurações adotadas pela UFRN?

b) Quem hospeda o servidor?

c) Por que virtualizar o servidor?

2. A integração do SEER – LOCKSS está efetivada de forma satisfatória?

3. Como é feito o monitoramento da caixa LOCKSS e dos processos de preservação?

a) Então o monitoramento é realizado apenas pela Coordenação da Cariniana?

4. Quanto ao espaço disponibilizado para armazenamento dos conteúdos digitais, na caixa LOCKSS da UFRN, considera satisfatório?

INVESTIGAÇÃO DE MESTRADO

Repositórios e preservação digital: proposta de requisitos para a integração do RI UFRN com a Rede Cariniana.

Mestranda | Elisângela Alves de Moura

Orientadora | Doutora Maria Cristina de Carvalho Alves Ribeiro

Coorientador | Doutor Miguel Ángel Márdero Arellano

GUIÃO DE ENTREVISTA

NOTA INTRODUTÓRIA:

- Apresentação individual

- Apresentação do estudo

À coordenação da Rede Cariniana

1. Sobre a Rede Cariniana, quando foi criada?

a) Há documento normativo instituindo a Rede?

2. De acordo com as apresentações (slides) disponibilizadas no Portal da Rede Cariniana, na primeira PLN constava a Universidade Estadual do Maranhão - UEMA. Porém, a instituição não consta na estrutura atual. A universidade saiu da Rede? Pode especificar os motivos?

a) O que ocorre caso uma instituição decida não mais participar da Rede Cariniana?

3. Ainda de acordo com as apresentações há estruturas que mostram PLN formada por região e outra por tipo de documento. Houve ou haverá subrede por região, ou somente por tipo de material?

4. Efetivamente já está sendo preservado outro tipo de material para além dos periódicos armazenados em SEER?

5. Há definição quanto ao fluxo de trabalho para materiais em DSpace, similar a disponibilizada para periódicos?

6. Para o SEER, fez-se necessária a incorporação do manifesto LOCKSS. Para o DSpace, será necessário fazer o mesmo?

7. Quanto as caixas LOCKSS, é possível visualizar apenas o conteúdo da caixa UFRN, ou de qualquer outro nó, individualmente?

8. Como será a identificação/visualização das teses e dissertações na caixa LOCKSS?

INVESTIGAÇÃO DE MESTRADO

Repositórios e preservação digital: proposta de requisitos para a integração do RI UFRN com a Rede Cariniana.

Mestranda | Elisângela Alves de Moura

Orientadora | Doutora Maria Cristina de Carvalho Alves Ribeiro

Coorientador | Doutor Miguel Ángel Márdero Arellano

GUIÃO DE ENTREVISTA

NOTA INTRODUTÓRIA:

- Apresentação individual e institucional
- Apresentação do estudo

Ao Técnico de TI da Rede Cariniana

1. De acordo com as especificações indicadas pela Rede Cariniana para instalação e configuração das caixas LOCKSS, uma das solicitações é: “Performance mínima aceitável: Core 2 Duo”. Contudo, a caixa LOCKSS do IBICT foi instalada com configuração inferior. O fato de ter processador de um ou dois núcleos compromete os processos?

2. A partir da interface administrativa da caixa LOCKSS UFRN, não é possível inserir um novo título de periódico (que tenha sido incorporado recentemente ao portal de Periódicos por exemplo). Como é realizado o registro de novos títulos para que possam ser ingeridos pelo LOCKSS, e preservados?

a) Sempre que um novo título de periódico for incorporado ao Portal da UFRN, por exemplo, terá que ser comunicado a Stanford para registro e posterior ingestão?

b) Como será o processo de ingestão para teses e dissertações? Terá que ser feito algum tipo de registro junto ao nó central, similar ao que é feito para periódicos?

3. A interface administrativa não apresenta nenhuma funcionalidade específica referente a teses e dissertações, DSpace ou qualquer outro tipo de material. Será incorporada alguma funcionalidade específica?

4. Com relação ao processo de ingestão de conteúdo pelo LOCKSS, qual a periodicidade dessa ingestão?

a) Não há necessidade de enviar o comando para ingestão manualmente?

5. Após a ingestão e a partir do armazenamento, os conteúdos são submetidos ao processo de *Poll*, principal responsável pela integridade e autenticidade dos dados. Com que regularidade esse processo ocorre?

APÊNDICE B – Fontes analisadas relativas ao contexto institucional de cooperação

FONTES ANALISADAS	
UFRN	
Política de Informação	http://repositorio.ufrn.br:8080/jspui/documentos/resolucao_592010_consepe_riufrn.pdf
Projeto de Criação RI UFRN	Documento interno UFRN
Site RI UFRN	http://repositorio.ufrn.br:8080/jspui/
REDE CARINIANA	
Acordo de Cooperação Técnica	http://cariniana.ibict.br/index.php/publicacoes/item/14-acordo-de-cooperacao-tecnica
Apresentações da Rede Cariniana (.ppt)	http://cariniana.ibict.br/index.php/2014-07-15-17-27-30
Guia da Interface Administrativa do LOCKSS	http://cariniana.ibict.br/index.php/publicacoes/item/1-manual-de-ial
Guia de Preservação Digital de Periódicos Eletrônicos na Rede Cariniana	http://cariniana.ibict.br/index.php/publicacoes/item/72-guia-de-preservacao-digital-de-periodicos-eletronicos-na-rede-cariniana
Manual de Instalação LOCKSS	Não consta online
Plano de Trabalho das Equipes da Rede Cariniana	http://cariniana.ibict.br/index.php/publicacoes/item/18-plano-de-trabalho-cariniana-2014-2018
Política da Rede Brasileira de Serviços de Preservação Digital	http://cariniana.ibict.br/index.php/publicacoes/item/27-politica-da-rede-brasileira-de-servicos-de-preservacao-digital
Portal da Rede Cariniana	http://cariniana.ibict.br/
Relatório Técnico: análise das caixas LOCKSS	http://cariniana.ibict.br/index.php/publicacoes/item/28-relatorio-tecnico-caixas-lockss

